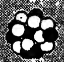
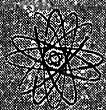
MAN HAS TAKEN MEASURE

Shown below is a scale of the dimensions of the universe as
atom's nucleus, as the universe's inside dimension, to the dis-

	ATOM'S NUCLEUS	$\frac{1}{1{,}000{,}000{,}000{,}000}$ CENTIMETER
	ATOM'S OUTSIDE DIAMETER	$\frac{1}{100{,}000{,}000}$ CENTIMETER
	ELECTRON MICROSCOPE'S LIMIT OF VISION	$\frac{1}{1{,}000{,}000}$ CENTIMETER
	FLU VIRUS	$\frac{12}{1{,}000{,}000}$ CENTIMETER
	TOBACCO MOSAIC VIRUS	$\frac{28}{1{,}000{,}000}$ CENTIMETER
	LIGHT MICROSCOPE'S LIMIT OF VISION	$\frac{1}{10{,}000}$ CENTIMETER
	RED BLOOD CELL	$\frac{4}{1{,}000}$ CENTIMETER
	HEAD OF A PIN	$\frac{1}{10}$ CENTIMETER
	CENTIMETER	
	INCH	

OF ATOMS AND GALAXIES

so far measured by man. It ranges from the diameter of an

tance to the farthest galaxy recorded by astronomers.

SIX-FOOT MAN	72	INCHES
NEW YORK TO SAN FRANCISCO	3,000	MILES
DIAMETER OF EARTH	8,000	MILES
DIAMETER OF SUN	864,000	MILES
EARTH'S ORBIT DIAMETER	186,000,000	MILES
ONE LIGHT YEAR	6,000,000,000,000	MILES
NEAREST STAR	25,000,000,000,000	MILES
DIAMETER OF OUR GALAXY	100,000	LIGHT YEARS
DISTANCE TO FARTHEST OBSERVED GALAXY	300,000,000	LIGHT YEARS

Courtesy of "Life" Magazine

Pneumococcus (×67,000) taken from the peritoneal cavity of a dead mouse.

The ELECTRON MICROSCOPE

By

E. F. BURTON
Head of the Department of Physics
and Director of the McLennan Laboratory
University of Toronto

and

W. H. KOHL
Development Engineer
Rogers Radio Tubes Limited
Toronto

Drawings by Dorothy Stone

REINHOLD PUBLISHING CORPORATION
330 WEST FORTY-SECOND STREET, NEW YORK, U. S. A.
1942

Printed in the United States of America by
INTERNATIONAL TEXTBOOK PRESS, SCRANTON, PA.

Preface

"I was received very kindly by the warden [of the Grand Academy of Lagodo], and went for many days to the academy. Every room hath in it one or more projectors [researchers] . . . .

"The first man I saw was of a meager aspect with sooty hands and face, his hair and beard long, ragged and singed in several places. His clothes, shirt, and skin were all of the same color. He had been *eight years upon a project for extracting sunbeams out of cucumbers,* which were to be put into vials hermetically sealed, and let out to warm the air in raw, inclement summers. He told me he did not doubt, in eight years more, he should be able to supply the governor's gardens with sunshine at a reasonable rate; but he complained that his stock was low, and entreated me to give him something as an encouragement to ingenuity, especially since this had been a very dear season for cucumbers. I made him a small present. . . . ." *Gulliver's Travels, Voyage to Laputa,* etc., Chapter V.

With the exception that sartorial and sanitary arrangements may have advanced with our civilization, there is an uncanny parallel between the "projector" of Lagodo in 1741 and the electron microscope investigators at Toronto in 1941. We have *"been eight years upon a project" for taking magnified pictures without light,* an investigation which is, to the modern Gulliver, no less foolish than the cucumber experiment was to the original Gulliver.

We "do not doubt, in eight years more," that there will be an electron microscope in every "project" laboratory and that it will contribute greatly to our knowledge of things as remote from each other as industry and medicine.

A third item in the parallelism is apparent, namely "that our stock is low," and we still entreat contributions "as an encouragement to ingenuity," especially since our "cucumbers" are still

expensive. In spite of the fact that the electron microscope bids fair to reveal to the eyes of man in the domain of the very minute things quite as wonderful as Galileo's telescope revealed of the wonders of space, we still have the feeling that it is ever "a dear season for cucumbers."

E. F. Burton
W. H. Kohl

Toronto, Ontario
March 1, 1942.

Acknowledgements

First in the list of acknowledgements, I wish to thank my co-author, Dr. W. H. Kohl, for the great help he gave the Department of Physics of the University of Toronto, at the beginning of the work on the electron microscope. It is on account of this assistance that Dr. Kohl was the natural person to call upon for co-operation in the preparation of the present book. For many years he has been a member of the staff of the Research Laboratory of the Rogers Radio Tubes Limited at Toronto, and for the greater part of this time he has served as special lecturer at the University.

His first course of lectures was on electron optics and, as he was engaged in development work on cathode ray tubes in an industrial laboratory, he was in a position to repeat some of Brüche and Johannson's experiments on the electrostatic electron microscope and produce images of oxide cathodes which revealed helpful information on the structure of the coating. Such an electron microscope was demonstrated before the seminar of the McLennan Laboratory in April, 1934; this probably was the first such demonstration in Canada. Images of oxide cathodes produced with magnetic lenses were demonstrated during a special course of lectures on electron optics in January, 1935.

Active research work was begun in this field at the Toronto laboratory during the fall term of 1935, by C. E. Hall, who constructed a simple electrostatic electron microscope and repeated quantitative measurements of Johannson on the properties of the immersion objective. In 1937 Mr. Hall joined the research laboratory of the Eastman Kodak Company, Rochester, New York, where he continued work in this field.

In the fall of 1937 J. Hillier, B.A. (Toronto), and A. Prebus, B.A. (Alberta), undertook the construction of a high-voltage

magnetic compound microscope with the aim of applying it to the investigation of biological specimens. Their work was highly successful and was first described in the *Canadian Journal of Research* in April 1939.

In February 1940, Mr. Hillier joined the laboratories of the R.C.A. Manufacturing Company and proceeded to design, in collaboration with Mr. A. W. Vance and under the direction of Dr. V. K. Zworykin, a type of electron microscope adapted to industrial research. Mr. A. Prebus joined the Ohio State University in July 1940. To Messrs. Hall, Hillier and Prebus we are greatly indebted, and also to Professor Arnold Pitt.

The work in Toronto is now being carried on by Mr. W. E. Ladd, Dr. L. T. Newman and Mr. J. H. L. Watson, who have assisted in the construction of a second instrument.

For financial assistance and encouragement, I wish to thank very heartily the following:

The National Research Council of Canada for direct grants in aid and scholarships.

The Banting Department of Medical Research of the University of Toronto, through the interest of the late Sir Frederick Banting and his staff.

The Banting Foundation for Medical Research for grant in aid toward work on viruses being carried out with the excellent cooperation of Dr. J. Craigie of the School of Hygiene, University of Toronto.

The Columbian Carbon Company of New York City and particularly Mr. W. B. Wiegand, their Director of Research, and the members of his staff, Mr. H. A. Braendle and Dr. Carl Sweitzer; this financial aid came at a time when it was needed most. It is pleasant to recall that all three are former students of the undersigned.

The Ontario Mining Association for a grant to the study of mine dust as related to silicosis. This work was undertaken through the interest of Dr. D. A. Irwin of the Banting-Best Medical Research Department.

The President and Board of Governors of the University for sympathetic assistance at all times.

The staff of our workshop and glass blowing plant, through whose ability and energy the successful construction was made possible. Two complete instruments were entirely produced in our own shops.

E. F. Burton

Contents

Chapter 1

Vision

The human eye is a wonderful optical instrument, more wonderful than any that man has ever devised. It unfolds to our mind the beauty of the world that surrounds us and is the means by which we react to and participate in it. But fascinating and manifold as is the scene which the naked eye presents, it does not satisfy our curiosity. We like to see things which the unaided eye cannot see, to bring far distant and ordinarily invisible stars within our view and to penetrate into the realm of the very small things, the microcosmos. For this purpose we need magnifying glasses: telescopes in the field of astronomy, and microscopes in the study of the very minute.

Whenever we interpose an artificial optical system between the eye and an object we are attempting merely to assist our eyes so as to widen their scope of vision. The eye is always the indispensable part of a sequence: object—vision aid—eye. In this book we are particularly concerned with the microscope. We must then keep in mind that it is only a link in this chain which we have just described.

Two almost self-evident conditions of vision must be recalled at the very beginning; first, an object is made visible by the light which it diffuses, scatters or reflects, and secondly, the position of the object viewed is judged to be along the direction of the line along which the light finally enters the eye. This latter circumstance lies at the basis of most optical illusions.

The Bending of Light Rays: Refraction

The following simple parlor trick illustrates quite well what happens to a beam of light as it passes from one transparent

medium to another. Place a coin on the bottom of a mug and see that the coin is well enough illuminated to be seen when viewed from above. Have an observer keep his head in such a position that he just *can't see* the coin as he peers over the edge of the mug (Fig. 1a). If while the observer keeps his eye fixed

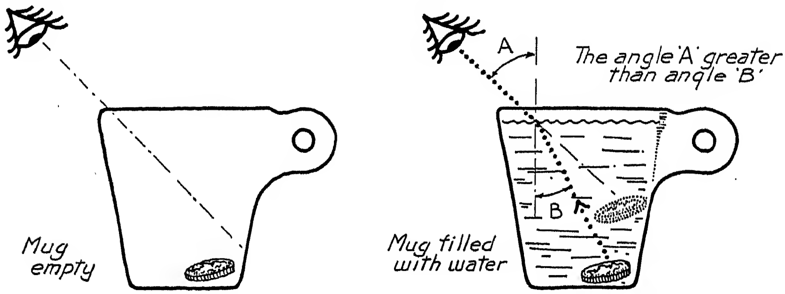


FIG. 1a. FIG. 1b.
How light bends as it passes from water to air.

in position, the mug is filled with water, the coin comes into view (Fig. 1b). The broken straight line represents the path along which the light appears to travel from the coin to the eye. But we know that the coin has not changed its position. Therefore the real path of the light from the coin to the eye must be that marked by the dotted line.

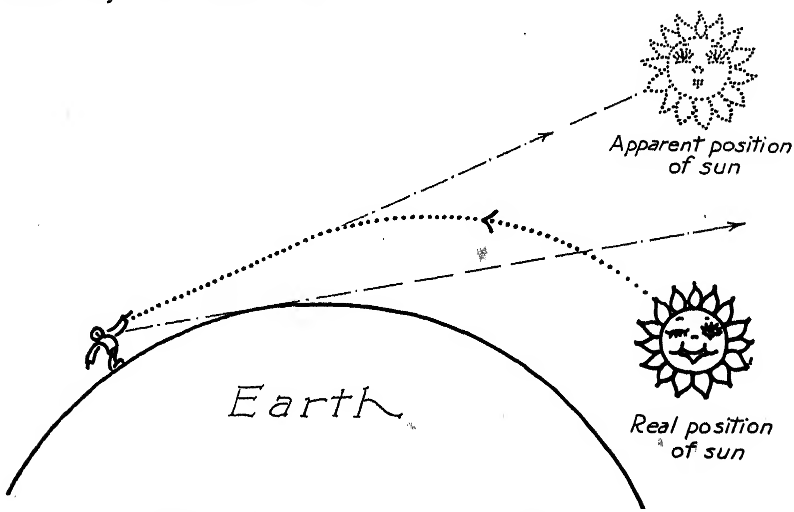


FIG. 2. How one can see the sun after it has set.

Another example of such an optical illusion, on a grand scale, is the fact that we see the sun at sunset for some considerable time after it has really sunk below the horizon. The reason for this is that the light from the sun is gradually bent around in going through layers of the atmosphere of differing densities as shown in Fig. 2. Our judgment of the position of the sun is dictated by the direction along which the light finally enters the eye.

These two phenomena are perhaps the earliest recorded observations having to do with light; they were described by Cleomedes in A.D. 50.

The Direction of Bending of a Light Beam

Very cursory examination of Fig. 1b tells us how the light bends as it passes from water to air: the angle which the emergent ray makes with the side of the mug, or rather with the perpendicular to the surface of the water, is *larger* than the angle which the ray of light from the coin makes with the side of the mug. In general, we may say that when light is going obliquely from a dense medium (water) into a less dense medium (air) it bends away farther from the normal to the surface of separation of the media.

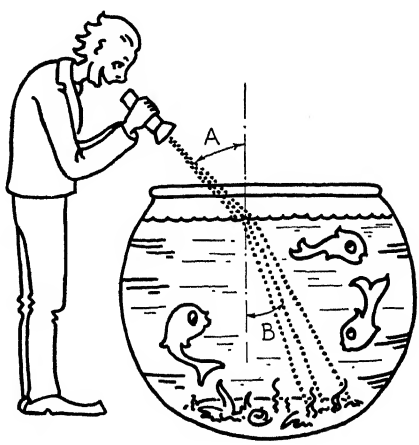


FIG. 3.

How light bends as it passes from air into water.

The angle 'B' is less than angle 'A'

Now we can easily perform the reverse experiment. If we direct a narrow beam of light, say from a flash-light, obliquely onto the surface of some slightly turbid water contained in a glass bowl, we can see easily that, as the light goes from the air (the less dense medium) into the water (the denser medium), the direction of the ray is bent nearer to the normal of the surface (Fig. 3). This change of direction which a light beam undergoes when it passes from one transparent medium into another is described by the term *refraction.*

The Prism and the Lens

We are all familiar with effects produced by blocks of glass or automobile lamp lenses in causing a change in the direction of light passing through them. Fundamentally all such effects can be simply illustrated by tracing a ray of light through a glass prism, which is a rather long uniform block of glass with a triangular cross-section (Fig. 4).

Consider a candle placed at the point, P, and one's eye at the point, E. If we interpose a glass prism, ABC, in our line of vision, EP, we would judge that the candle is displaced to the position Q. Light travelling along the line PE will now be turned off the track at K and will not reach the eye. The ray which

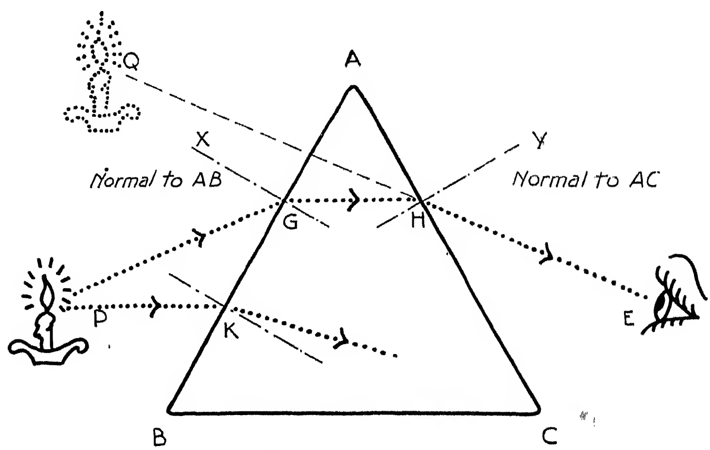


FIG. 4. How a light beam is deviated by a prism.

does reach the eye is one travelling from the candle in some such direction as PG. At the point G, as the light is passing from a rare medium (air) to a dense medium (glass) its direction inside makes an angle with the normal to the surface at the point G, which will be less than the angle XGP; and then as it passes out of the glass again into the air the angle YHE will be larger than that which the ray, GH, makes with the normal to AC. The eye then judges the candle to be along the direction EHQ, which is the direction of the straight line along which the light enters the eye. The prism has thus caused a distinct change in the direction of the ray which ultimately enters the eye.

After this account of the deviation of a beam of light caused by a prism, it is easy to understand the action of a simple lens, like a reading lens, when it is used to "focus" on a screen the image of a distant source such as an incandescent light or a brightly illuminated window frame.

It is a matter of quite common practice to interpose such a lens between a bright object, (O, Fig. 5a), and a screen, e.g., one's hand, and to produce on such a screen an image such as at I. One has always to adjust the position of the lens and the screen in order to obtain the position of I where the image is sharp. If the experiment is performed in a darkened room and the source is shielded so as to send out light only in the direction of the

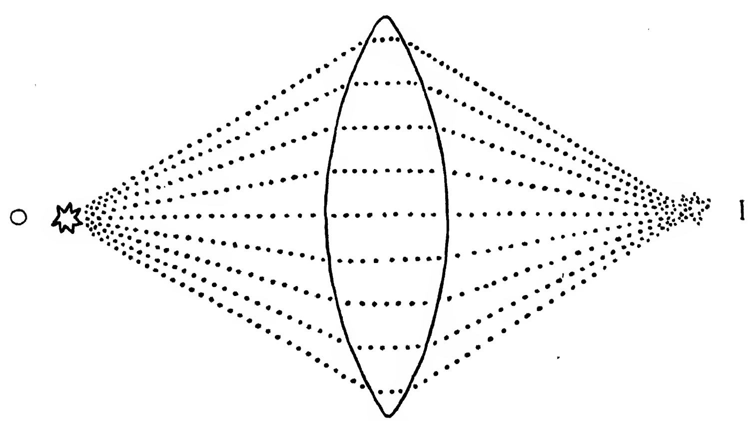


FIG. 5a. How a lens bends light.

lens, the actual path of the light from O to I can be seen, particularly if some smoke or chalk dust is blown into the region round the lens.

The reproduction at I is called the *image* of the object, and the process of obtaining a sharp image is spoken of as *focussing* the lens.

The explanation of this action of the lens is quite clear when we realize that any lens can be considered as being built up of a great number of portions of prisms, as indicated in Fig. 5b. By following the paths of rays from O through the various prism sections, it is apparent that it is quite possible for the emergent rays all to pass through the point I.

It is in the action of such lenses that we find the explanation of the functioning of the eye and of the role played by the microscope as an aid to vision.

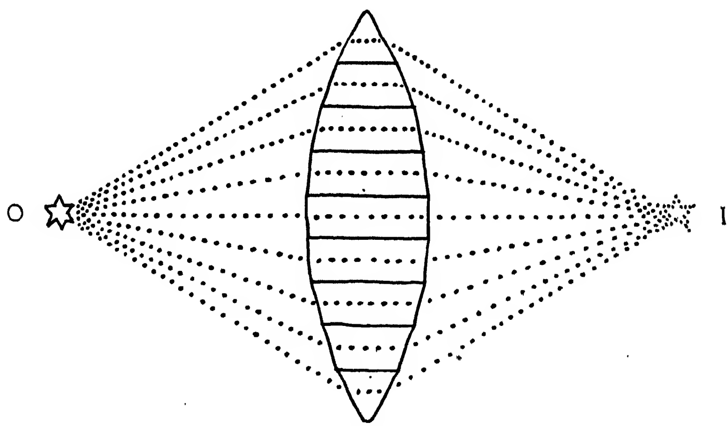


FIG. 5b. How a lens may be considered as made up of sections of prisms.

The Eye

The eye is nature's camera, the prototype of all cameras. The structure and function of the various parts of the eye can be very well understood by referring to the corresponding parts of an ordinary camera.

The simplest possible form of camera is the so-called pin-hole camera—a light-tight box with a pin hole in the center of one side

and a photographic plate or film placed against the inner side directly opposite the hole (Fig. 6).

If the box is turned with the pin hole toward an object, such as the animal at AD, and the hole is uncovered so as to let light into the box, an image of the animal will be impressed on the plate or film. This image will be produced in the following manner. Each part of the animal that faces the camera sends light into the pin hole, the aperture. The amount of light which is diffused by any portion of the surface of the animal depends on the natural color of the surface. Practically no light will come from black

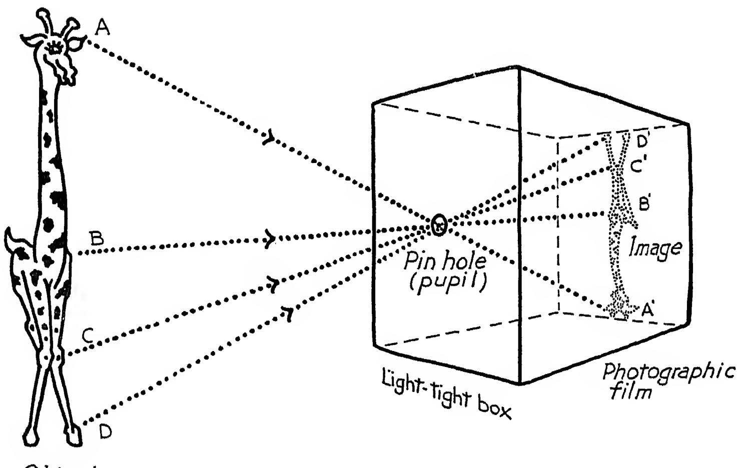


FIG. 6. A pin-hole camera.

patches and a great deal of light from white patches. Since light travels in straight lines from any surface point on the animal to the aperture, all rays will cross one another in the aperture and continue on to the photographic plate. Thus the light ray coming from a point A on the head of the animal will impinge on the film at point A′ and blacken the film according to its intensity. Thus we will have the distribution of light and shade at the surface of the object accurately reproduced on the film and obtain a picture of the animal on developing the film.

The pin-hole camera, though interesting as a hobby and instructive as a scientific experiment, has its limitations. The

amount of light entering the pin hole is so small that a long exposure is always necessary and so the use of this camera is limited to still objects. As the distance from the pin hole to the film is fixed there is no means of adjusting for variable focus, and consequently no one part of the picture is more sharply in focus than any other.

We now turn to the simple camera, which is familiar to all (Fig. 7). This consists of a light-tight box provided with a convergent lens through which the light enters, and a plate or film on which the image formed by the lens can be focussed. As

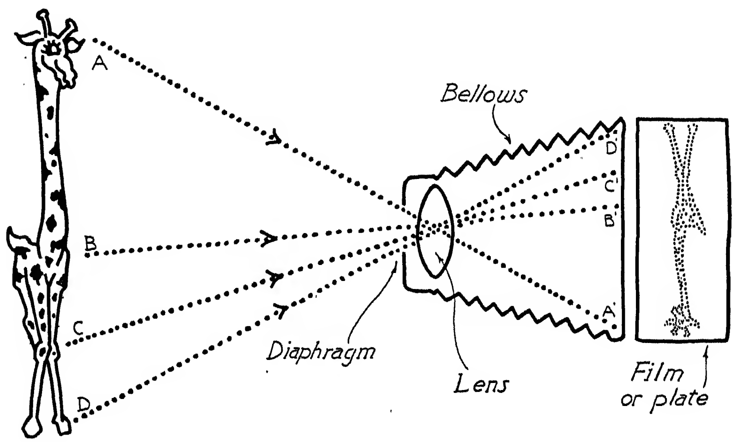


FIG. 7. An ordinary camera.

referred to above (Fig. 5a), when we wish to focus an image made by a lens on any surface the relative distances between the object, the lens, and the image screen must be adjustable. This is provided for in the camera by having the light-tight box furnished with a bellows body so that the lens-film distance may be adjusted without letting in any extraneous light.

The advantage of such a camera over the pin-hole type is that more light from any point on the object is concentrated at the corresponding point on the image, and therefore a much smaller exposure is necessary in order to cause a definite effect on the film. It is consequently possible to take 'instantaneous' exposures of moving objects.

The functioning of the human eye may now be readily understood, at least in its basic principles, from what we have said about the working of a camera.

The eye-ball is a light-tight spherical box with a great part of the inner surface covered with a sensitive film, called the retina, from which light impressions are transferred to the brain (Fig. 8). Just as in the case of the camera, the relation between the positions of the lens, the object and the image must be adjustable in order to obtain a sharp image on the light-sensitive film (the retina). With the eye, however, these adjustments are not made

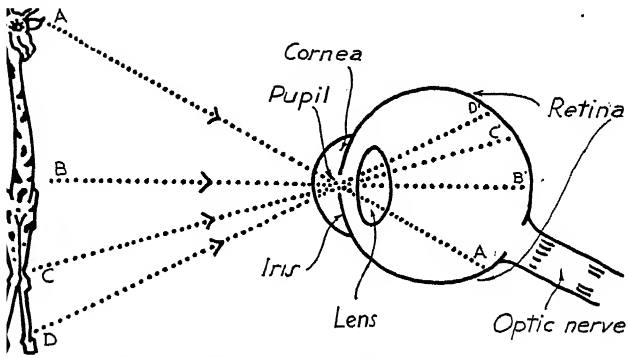


Fig. 8. How the eye forms an image.

by changing the *position* of the lens but by individually adjusting the size and shape of the lens itself, so that the image is always formed on the retina, which remains at a fixed distance from the pupil of the eye. That is, if we wish to look at a distant object we automatically adjust the eye lens so that the image of the distant object will be focussed on the retina; as the object comes nearer and nearer, the eye lens is adjusted through muscular control until one sees the object distinctly in any new position. For a normal person the arrangement of the muscles regulating the eye lens gives the greatest comfort when the eye is allowed to rove over distant objects and no effort is exerted to bring any particular thing into focus. When one concentrates on carefully viewing near objects, the eye tends to become tired and suffers from

strain. It is impossible for the normal eye to focus on objects nearer than about 10 inches (25 cms.). This distance is known as the least distance of distinct vision and represents a very important natural limit of vision. No matter how we try to help the eye by artificial aids, say by a microscope, we must end up with at least the illusion that we are observing the objects or images as though they were placed at a distance of 10 inches from the eye.

The Scope and Limitation of the Eye

The normal eye has a very wide field of vision and a great range in distance. It will be apparent from Fig. 9 that, as the

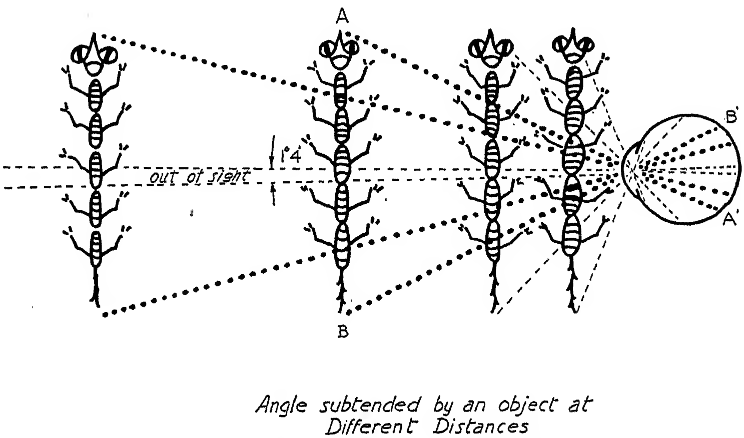


FIG. 9. How the size of the image on the retina decreases as the object is moved farther and farther from the eye.

object recedes from the eye, the area of the retina covered by the image, *e.g.*, at A′ B′, becomes smaller and smaller, and finally, as the distance increases very greatly, we say the object disappears. Light does not cease to come from the object, but the area of the retina affected by the light coming from the object becomes so small that, for some physiological reason, it ceases to produce a sensation of vision in the brain. It has been found that, if the angle subtended at the eye by the object becomes less than about 1.4 minutes (about two one-hundreths of a degree), the object ceases to be distinguishable.

From this it follows that whether any particular object is visible to the eye or not depends on *both its size and its distance* from the eye; in any case the visibility is limited by the angle subtended at the eye (see Table 1). Of course, we take for granted that the object is always brightly enough illuminated to make vision possible.

Table 1. Sizes of Objects Just Visible to the Eye at Different Distances.

Distance of Object from the Eye	Linear Dimension of Smallest Object Visible
10 miles	20 feet
1 mile	2 feet
100 yards	1 inch
10 feet	1/20 inch
10 inches	1/250 inch, 1/100 cm., 4 mils or 1/10 mm.

Now an object may be so small that it is invisible at a distance of 10 inches from the eye; this is true if it is so small that it subtends an angle at the eye of less than 1.4 minutes. Since the eye cannot focus on anything which is closer to it than 10 inches, we must conclude that the unaided ("naked") eye cannot see any object less than 1/250th of an inch in diameter, nor can it make out any detail of any marking on a large visible object if the linear dimension of the pattern is less than 1/250th of an inch.

In this extremity we call on the *microscope* to assist the eye.

Chapter 2

Light Microscopes

Aids to Defective Vision

Everyone is fairly familiar with the simplest cases of defective vision, such as short-sightedness and long-sightedness, and the means used to correct such defects. In one case, the eye is able to focus the image distinctly on the retina only when the object is held abnormally near the eye; in the other case, the object must be held abnormally far from the eye. If the position of the object is at the normal distance for distinct vision, *i.e.*, ten inches, then for such defective eyes the image is formed either in front of the retina or behind it. In either case the object is not seen distinctly (see Fig. 10).

These defects are corrected by the use of eye-glasses (or spectacles); for the case illustrated in Fig. 10a the eye-glass throws

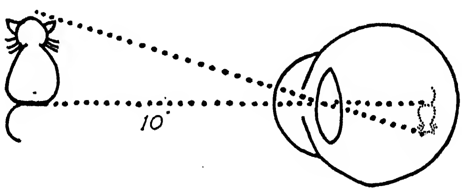


FIG. 10a. Near-sightedness.

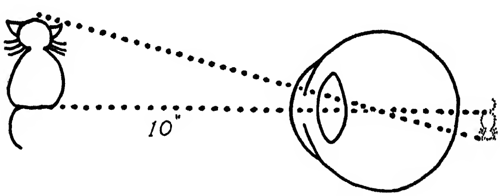


FIG. 10b. Far-sightedness.

the image farther away from the lens of the eye on to the retina (Fig. 11a), while, for that in Fig. 10b, the convergence of the light is increased as it enters the lens of the eye and thus the image is brought back to the retina (Fig. 11b).

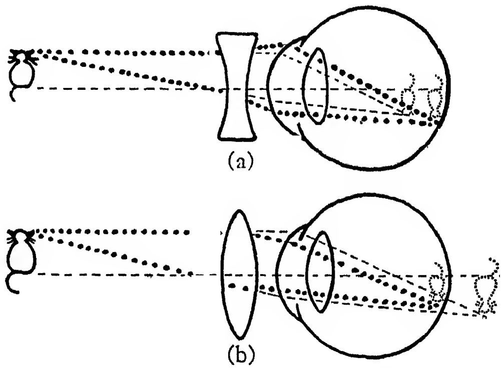


FIG. 11.

How glasses correct near- and far-sightedness.

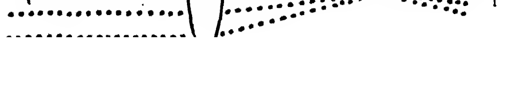
How to Tell the Position of the Image Formed by a Single Lens

Before explaining further the operation of these aids to vision, we shall have to present a few simple rules which enable us to estimate by means of a diagram the position of images formed by any lens. Simple lenses are of two general classes, converging and diverging, depending on what happens to a beam of light falling on the lens. This can be illustrated experimentally by observing what happens to a beam of sunlight, say, admitted through a hole in a window blind, when the beam falls on the lens. A simple application of the principles described in the first chapter regarding the bending of the light ray as it goes from dense to rare or from rare to dense media will justify the representation given in Figs. 12a and 12b.

In the case of the converging lens (Fig. 12a) the rays parallel to the axial line XY all pass through a point F on the axis; this point

FIG. 12a.

Converging Lens.



is called the *focus* of the lens. If the beam of parallel rays were sent through the lens in the opposite direction, from Y to X, there would be found a similar point, F′, on the opposite side of the lens; this is also called a focus. If the lens is symmetrically shaped on the two sides, the distance of F from the center of the lens will be the same as the distance of F′ from the center; this distance is called the *focal length* of the lens.

Rule 1: *Every ray entering the lens parallel to the axis XY passes out through the focus F, or F′, as the case may be.*

Using a beam of sunlight if a screen or the palm of the hand is held at the focus F, or F′, a sharp point of light will be seen—in reality an image of the hole in the window blind.

If a diverging lens is used no such image will be formed on a screen placed anywhere; but, if the eye is placed in the position represented in Figure 12b, the observer will have the illusion that

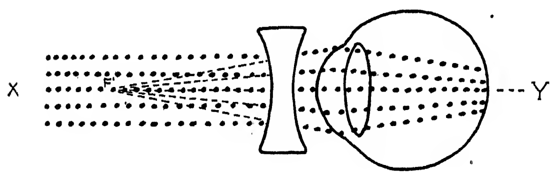


FIG. 12b.
Diverging Lens.

the light seems to come from a point F′ on the side of the lens remote from the eye. This point F′ is called the focus of the diverging lens; so for this case Rule 1 becomes: *Every ray entering the lens parallel to the axis XY passes out in a direction as though it proceeded from the point F′, the focus of the lens.* As in the case of the converging lens, if the diverging lens is symmetrical as to its two sides there will also be a similar focal point on the other side of the lens at the same distance from the center of the lens as F′.

There is a second rule applicable to both kinds of lenses. Rule 2: *Every ray entering the lens at any inclination and passing through the center point of the lens will emerge with no perceptible deviation from its original direction* (Fig. 13). This is strictly true only for very thin lenses, but it is permissible to use the rule when one is just approximating to the real conditions.

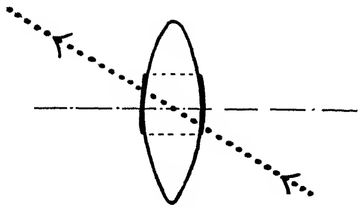


FIG. 13a.

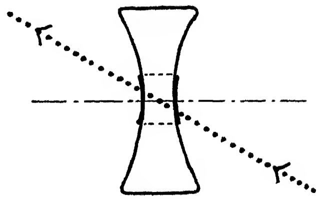


FIG. 13b.

For a thin lens the ray through the central point is not deviated perceptibly.

The two rules given above enable us to find the position of the image of any object very easily (Fig. 14). The image of A will be on the line AO produced through the lens and also along PF produced; these lines meet in the point A′ which is consequently the image of the point A. Similar procedure gives B′ as the image of B; consequently the figure A′ B′ marks the image of AB.

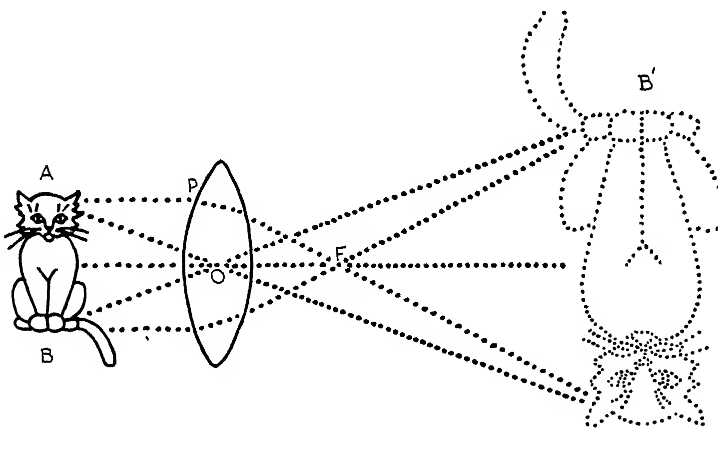


FIG. 14. How to determine graphically the position of the image of an object when using a single converging lens.

The Simple Microscope, Reading Lens or Magnifying Glass

It has been pointed out that even a normal eye is unable to form an image of the object on the retina, or in other words cannot see the object, if it is moved nearer to the object than ten inches. It is this circumstance that limits the magnifying power

of the eye. But here again we can call to our aid a simple converging lens. The eye lens itself is a single converging lens, and in Fig. 15 we show what happens to the image of an object as the object is brought closer and closer to such a lens.

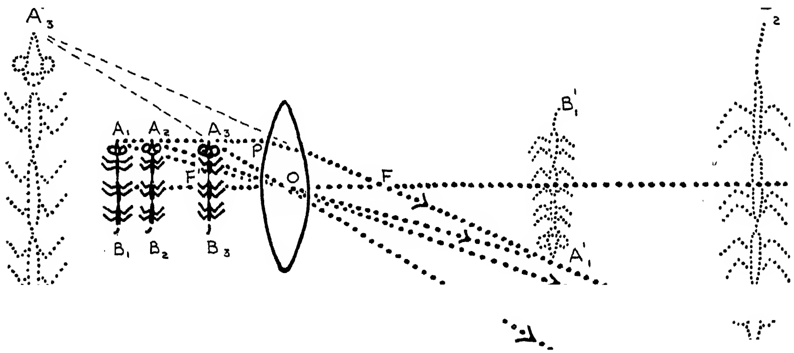


FIG. 15. How the position of the image changes as the object is moved closer and closer to a converging lens. The Simple Microscope.

The ordinary rules give the positions of the images of A_1B_1 and A_2B_2 at $A'_1B'_1$ and $A'_2B'_2$ respectively. However, when we follow the two standard rays from the point A_3 of the object A_3B_3 the light emerges from the lens along the lines PF and A_3O, which never meet to the right of the lens; but if the eye is placed to the right of the lens and close to it the observer will have the impression that the light appears to come from the position $A'_3B'_3$. Thus to the eye the object is enlarged and the position of the image must be at least ten inches from the eye in order to be in focus. This shows why we have to adjust the relative positions of the eye, the lens and the object in using such a simple microscope. It is at once apparent that the object must be placed between the focus of the lens and the lens itself and that the eye has to be very near the lens in order to gather in the diverging rays A_3O and PF.

The Compound Microscope

The ordinary form of a compound microscope merely uses two individual lens systems to produce magnification in two stages.

There are really two fundamental forms of the compound microscope; the first is represented by a combination of two separate systems, each similar to that shown in Fig. 14, and the second by a combination of two separate systems, as shown by combining Figs. 14 and 15.

For example, we may produce an enlarged image, A′B′, of an object, AB, by a simple lens (or single lens system) as in Fig. 14 and then, using this image as a new object, produce a still larger image by a second lens, as illustrated in Fig. 16. This new image,

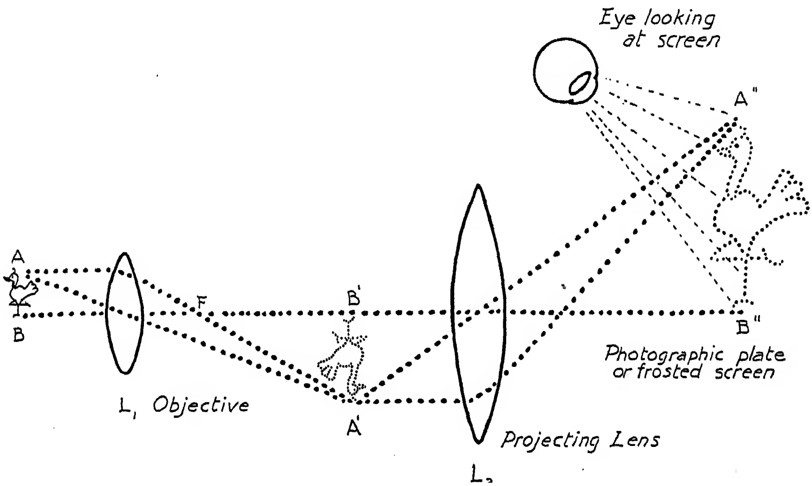


FIG. 16. A compound microscope as used to project or photograph the image.

A″B″, can be thrown on a screen of frosted glass and examined by the eye; it can also be projected onto a photographic film or plate; thus we may obtain a permanent record of the enlargement. The first lens or system, L_1, is known as the *objective* and the second lens system, L_2, as the *projector*. It is apparent that the second image, A″B″, can be treated again as a new object, and a still larger image may thus be obtained by another projector lens system.

The common usage of the term *compound microscope* refers to an instrument used directly by the eye. One looks into the eye-piece at the top of the instrument and sees the image. Fig.

17 shows diagrammatically the lay-out of the component lenses. The objective, L_1, forms the first image, A′B′, as shown in Fig. 16. This image, A′B′, now serves as the object for the eye-piece which acts merely as a simple microscope. Consequently, the right-hand portion of Fig. 17 is merely a repetition of Fig. 15; and, as described under the section on the simple microscope, the eye gets the impression that the light is proceeding from the image, A″B″, at a distance of ten inches from the eye.

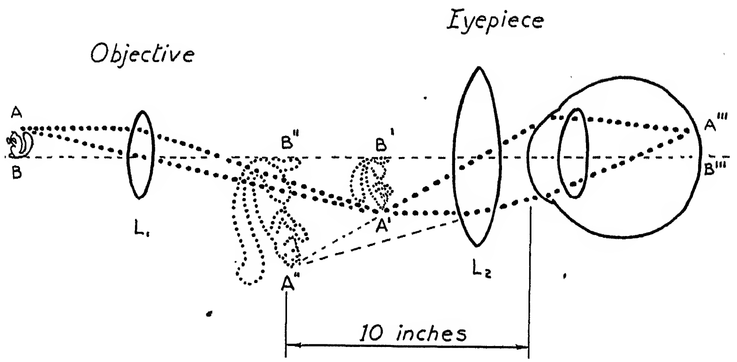


FIG. 17. An ordinary compound microscope for visual observation.

The lens of the eye gathers in the slightly diverging beam which leaves the eye-piece and focusses this beam on the retina. Of course, this same eye-piece might have been used on the image, A″B″, of Fig. 16, and the eye would have seen a final image of much greater magnification.

In fact theoretically there is no limit to the magnification that one might attain by using lens after lens after lens. However, in practice, there are limitations in construction and lighting which make a progressive magnification, ad infinitum, impossible. Some of these difficulties are noted in the next section.

Aberrations of Lenses and Lens Systems

In the preceding discussion we have tacitly assumed that all the lenses acted in an ideal fashion, that is, that the lens produced in the image a true representation of the object, point for point.

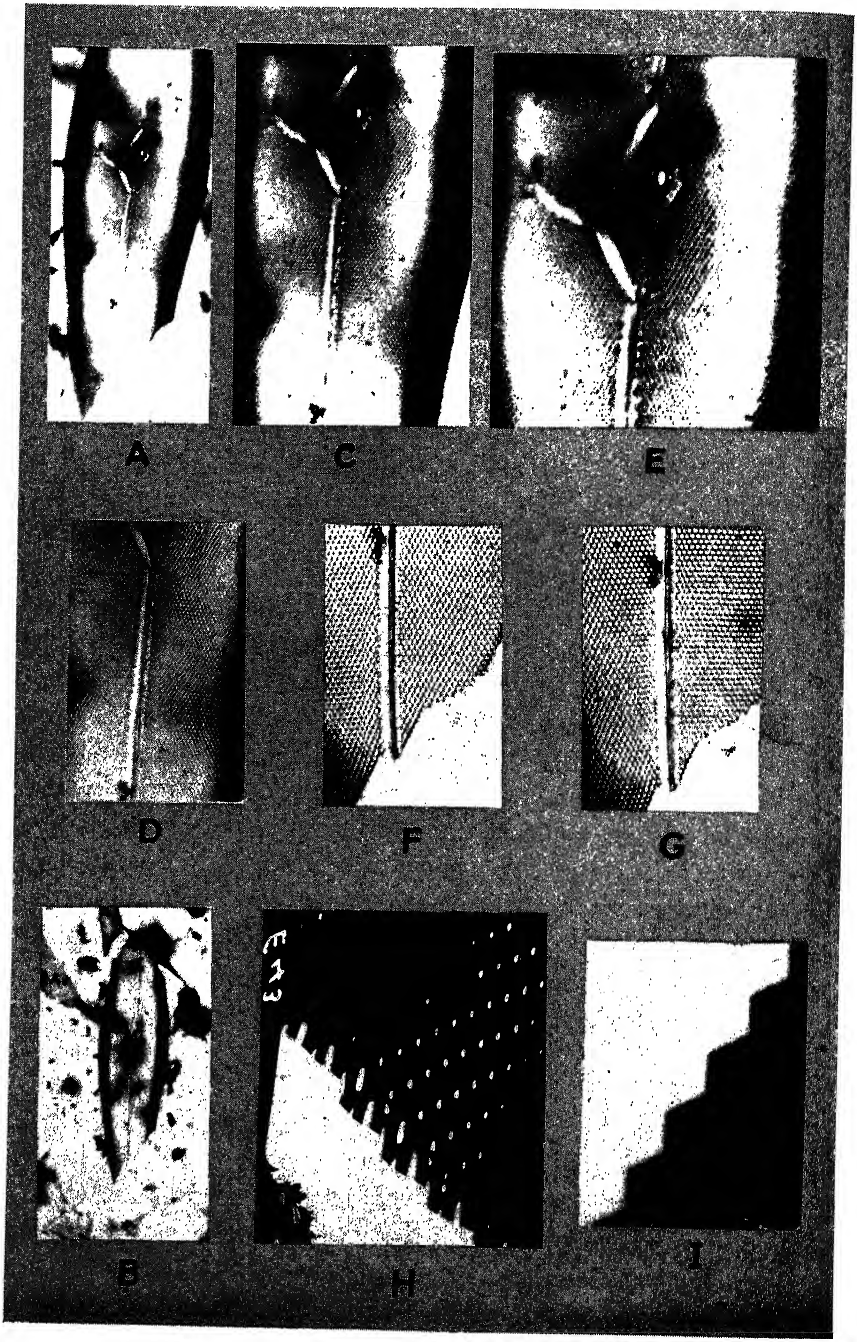
But there are many ways in which actual practice falls far short of this ideal. These shortcomings are known by the general name, *aberrations.*

If we look back at Figs. 5a and 14 we will find that we have assumed that light emanating from an object point is sharply focussed at a definite corresponding point in the image. But this is not so if we use only one lens. For one thing, the outer annular zone of the lens does not cause the light from a point to converge to exactly the same point as does the central zone; this defect of the lens is known as *spherical aberration.* Again, if we use ordinary white light, which as we know is just a mixture of light of different colors, we find that the different colors do not all converge to the same point. As a result the images are fringed with color—a defect known as *chromatic aberration.*

These are only two of the shortcomings of lenses. In order to overcome such defects, it is necessary to use a system of several individual lenses of differing shapes and made of different kinds of glass, instead of one single lens, as indicated in our figures. It would be quite beyond the scope of this book to deal at all fully with the subject of aberrations and their corrections. In fact, no matter how elaborate the system of lenses which makes up the objective or the eye-piece of an ordinary compound microscope may be, we may treat each system as equivalent to a single converging lens as far as our present needs require. But we must remember that striving for perfection in the reproduction of an object introduces complexity into the optical system *in theory* as well as *in practice;* attempts to produce a theoretically perfect microscope add greatly to the cost of a fine instrument.

The Resolving Power of a Microscope

The purpose of any microscope may be looked upon as *either* to enable us to see objects so small that they cannot be distinguished by the naked eye *or* to show us the finest detail in a large object. These two features really amount to the same thing, because if a small object, (Fig. 18), is visible in a microscope it is because two points, which are separated by a distance equal


A
C
E
D
F
G
B
H
I

Explanation of plate on page 30

The plate is arranged to show the relation of magnification, or enlargement, to resolving power, which depends on the numerical aperture (N.A.) of a microscope. All the pictures are of portions of the shell of a species of diatom, *Pleurosigma angulatum.*

A to G are optical (light) microscope pictures.

H and I are electron microscope pictures.

B is taken with N.A. equal 0.5, magnification $\times$100.

A, C, and E were each taken with the N.A. = 0.75 but with magnifications $\times$240, $\times$480, and $\times$960, respectively; there is not much increase in detail visible as the magnification increases.

D, F, and G were taken with progressively greater N.A., *viz.*, 0.85, 1.00, and 1.25, respectively, and with magnifications: $\times$420, $\times$900, and $\times$900. The detail is progressively better as the N.A. is increased.

H is an electron microscope picture with magnification about $\times$5,000. This shows the real position of the holes in the shell. I is a broken edge of the shell of about the same magnification as H.

The optical microscope pictures were taken by Dr. D. H. Hamly of the Department of Botany, University of Toronto.

to A-B on the object, can be distinguished as *separate* points. We have seen that for the unaided eye this limiting distance is about 1/250th of an inch. The limiting distance for any given microscope is known as its *resolving power*.

FIG. 18.

Separation of points and linear dimensions of objects as involved in discussion of resolving power.

Whether any two neighboring points on an object can be seen separated in the image given by any compound microscope depends on the objective alone. If they are not separated in the image formed by the objective, no amount of additional magnification by a projecting lens or by eye-pieces will ever succeed in separating them. This is illustrated in the series of photographs in the plate on page 30. Fig. A is the photograph of the image of a diatom shell with an initial magnification $\times 240$; Figs. C and E are successive enlargements of A, *i.e.*, $\times 480$ and $\times 960$ respectively. It is quite apparent that little really new comes out with the successive enlargements. On the other hand if different objectives are used, which give successively better and better resolving power, we obtain the results shown on page 30, Figs. D, F and G. The actual magnifications in the latter series of figures are $\times 420$, $\times 900$ and $\times 900$ respectively. Fig. B has a magnification about $\times 100$, and the smallest resolving power.

Theoretically, the question is, "What is the closest distance by which two points can be separated and still appear as separate in the image?" Practically we may ask, "How must the optical system be designed to bring about this separation in the image?"

Light Treated as a Wave Motion

In order to answer the last question it is necessary for us to speak of the nature of light. Up to the present we have been content to represent rays of light by means of straight lines and to follow the direction of the light by the path indicated by these

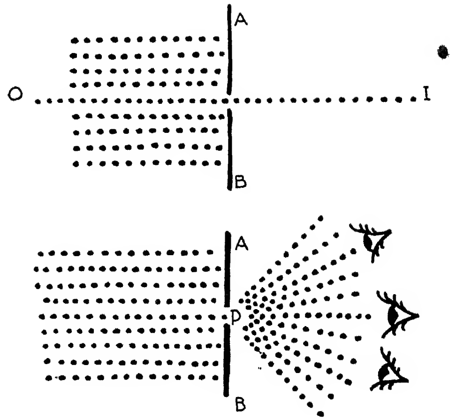


Fig. 19.

How light spreads on passing through a small hole in a screen.

lines or rays. In other words we have tacitly assumed that light travels along straight lines in any homogeneous medium as though it consisted of luminous particles shot out from the source. This theory of light is what is known as the *corpuscular theory*. If we followed this reasoning, we would judge that there could be no limit to the minuteness of detail in any object which we might be able to resolve and magnify.

We have already applied the principle of the corpuscular theory in the presentation of the simplest image formation by means of a single lens (Fig. 5a). In this figure we represent an ideal point, O, on an object as corresponding to an ideal point, I, on the image. According to this representation of light rays, if we should let a wide beam of light fall against a screen (AB, Fig. 19) which has a fine hole, P, we should expect a very thin straight and narrow beam of light to pass through P to I. What we really would observe, however, is that the very small amount of light which gets through the hole, P, spreads out in a fan shape (or really in the shape of a cone of light); and, if the eye is placed anywhere in the region to the right and directed toward the screen, some of the light that gets through P will enter it, and the eye will see the light coming from O.

This is an example of the phenomenon called *diffraction*, which is found experimentally to be a constant accompaniment of a wave motion. The experiment depicted above in Fig. 19

can easily be duplicated with water waves. If the bottom of a large flat pan, Fig. 20, is just covered with water and two flat metal bars are placed so as to make a barrier, AB, across the water surface with the exception of a narrow opening at P, the experiment corresponding to Fig. 19 for light can be performed by means of waves upon the surface of the water. Drop a little pebble into the water, say at X; the wavelets will spread out from X as a center and will be reflected back from AB *except* at the opening, P, where some of the motion gets through to the other side. But instead of passing along a narrow path as indicated by the line PQ, the motion that gets through P spreads out all around the surface in the form of little wavelets which seem to emanate from the point P as center.

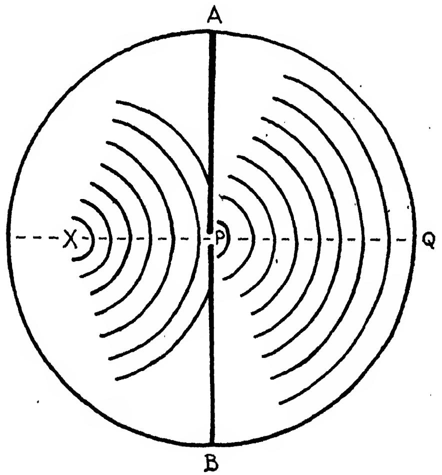


FIG. 20.

How a wave motion spreads on the surface of water after passing through a small opening in a bounding wall.

It is on account of analogies such as this that scientists were constrained to say that light acted as though it had the properties of a *wave motion* in some medium.

Once this phenomenon of diffraction of light was discovered, it was found that it had a very important role in the formation of images by lenses. If a small hole in a screen is illuminated from the left (Fig. 21) and used as an object point for a lens, L, and one uses just a small central portion of the lens, the image formed at the point, I, on a second screen will not be merely a sharp point

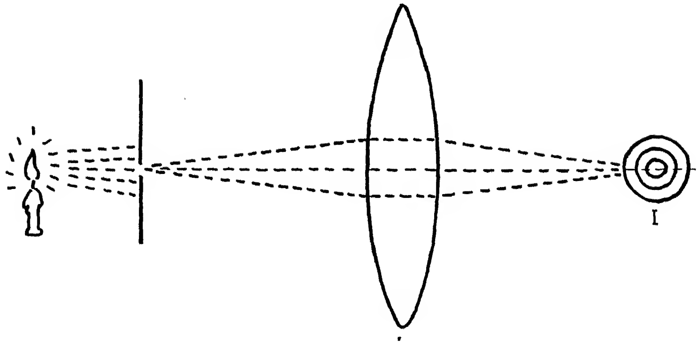


FIG. 21. The spurious disk or anti-point.

of light but a regular pattern as shown—a bright central disk surrounded by alternate dark and light rings. If one uses a telescope to view the image given by the lens, L, quite a number of such rings can be observed and easily photographed. Such a point image came to be known to microscopists as a *spurious disk* or *anti-point*. By no amount of manipulation can one get rid of this departure from an ideal image point.

If, instead of considering the object point as a bright point of light, we look at a small particle through a system of lenses, we observe in the image the converse of the image shown on the screen in Fig. 21, that is, a dark central disk surrounded by a series of alternate bright and dark bands. Now if we imagine a second small particle at O′ very near to the original particle at O, and just above it, the image pattern of the two points is as depicted in Fig. 22. The spurious disk images may overlap to such an

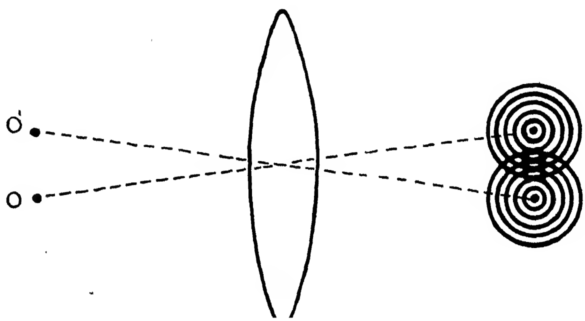


FIG. 22. Overlapping of the images of two neighboring object points.

extent that it becomes difficult to determine that there really are *two particles* in view. In such a case we say that the two particles in the object are not resolved. The critical problem is, "*How close together can O and O′ be and still have the lens system separate them for us in the image?*"

The late Lord Rayleigh solved this problem in a rather simple way, but before outlining his method we shall have to say a word or two about the interaction of two overlapping wave motions.

The Interference of Waves

Our whole conception of the interaction of wave motions comes from our observation of ripples or wavelets on the surface of liquids, such as caused by dropping a stone in water. As the stone strikes the surface, ripples spread out from the point of impact as center (Fig. 23a). If we could make an instantaneous

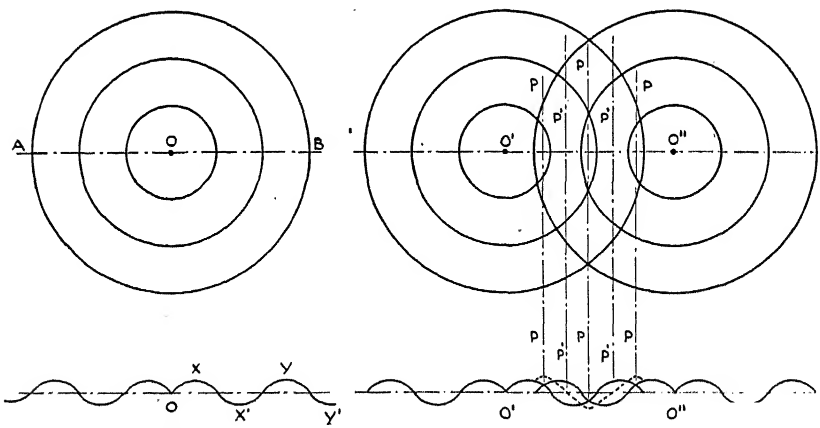


Fig. 23a. Fig. 23b.

The superposition of two systems of waves on the surface of water.

photograph of a section of the rippling surface in a plane perpendicular to the surface and through the line AOB, the ups and downs would be represented by the uniform wavy line in Fig. 23a. The distance, xy (or $x'y'$), from crest to crest (or trough to trough) is known as the *wave length*. Now suppose that two stones are dropped in at about the same instant at points

O′ and O″. Two sets of wavelets would develop from the centers O′ and O″. These two would overlap as they proceed; Fig. 23b represents an instantaneous picture of a section of the surface through the points O′ and O″. The surface particles at any point, P, will have assumed the displacement due to the sum of the two motions, and so the real position of the surface points at the particular instant chosen will be that given by the dotted curve. At the position P′, for example, the surface point will be at its original position of rest; that is, at this point the two wave motions combine to give zero displacement. Again, at the position P, there will be an abnormally large displacement due to the two disturbances adding their effects at this particular instant.

This interaction of two independent wave motions is known technically as interference.

From the above example we see how it is possible, by looking upon light as a wave motion, to realize that two beams of light may conspire to give zero light (darkness) at some points and excessive brightness at other points at the same instant. This view offers a rational explanation of the bright and dark rings present in the spurious disc in the image of a small point-source of light (or a small particle) produced by a microscope.

This phenomenon of interference forms the basis of Lord Rayleigh's determination of the resolving power of a lens or lens system such as a microscope objective.

The Limit of Resolution of a Microscope

We may represent the lens system by means of an ideal single lens (Fig. 24). The lens L′L″, represents diagrammatically the objective of a microscope. The center of the image of A will be at A′, the center of the image of B, at B′. *What is the shortest distance by which A and B can be separated in an object and still allow A′ and B′ to be distinguishable as two distinct points in the image?* Lord Rayleigh's argument is that A′ can be distinguished from B′ if the spurious discs belonging to the centers A′ and B′ do not overlap too much. This amounts to saying that A′ and B′ will be distinguishable if the part of the light wave motion which comes

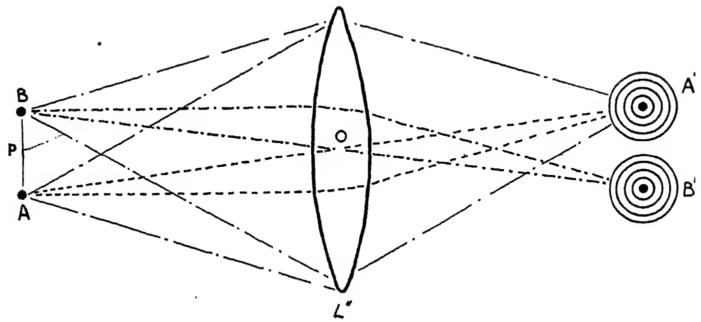


FIG. 24. Lord Rayleigh's determination of the resolving power of a lens system. The angle A in the text is one-half of the angle L′ P L″, if P is the mid-point of AB. To fit the text exactly AOA′ should be the axis.

from B and which arrives at the point A′ does not spoil the effect at A′ due to the light wave motion coming from A itself.

Whether the resultant effect at A′ due to disturbance from B is large or small depends upon the difference of length between the paths, BL′A′ and BL″A′, or in effect, the difference between BL′ and BL″, since, as L′A′ is equal to L″A′, the lengths of the paths in the region to the right of the lens are equal. Lord Rayleigh showed that A′ and B′ would be distinguishable if the difference BL″–BL′ were at least equal to the wave length of the light used. It can be shown rather easily that the smallest value which A-B could have, say *d*, would be given by the expression

$$d=\lambda/2 \sin A$$

where λ is the wave length of the light in the space between AB and L′L″, *i.e.*, in the region where the difference of path BL″–BL′ comes in, and sin A is the ratio L′O/L′P fixed by the small angle indicated in Fig. 24. The expression for *d* may also be written in the form

$$d=\lambda_0/2n \sin A \qquad \text{or} \quad d=\lambda_0/2\text{N. A.}$$

where λ_0 is the wave length of the light when in a vacuum, and *n* is the index of refraction of the medium between AB and the lens, L′L″. The angle A is called the *aperture* of the lens and *n* sin A, usually printed N.A., is called the *numerical aperture*.

According to this formula, the power to distinguish two points which are very close together in an object is limited in practice by the value of the wave length of the light used and by the value of the symbol N.A. The smaller the wave length of the light used, the smaller the limiting distance AB may be; the larger the numerical aperture, N.A., the closer together the points A and B may be. In the plate on page 30 are examples of microphotographs of the same diatom with the following values of N.A.: B =.5; A, C and E=.75; D=.85; F=1.00; G=1.25. It is quite apparent that detail is progressively better as the value of N.A. increases.

The maximum value of N.A. for any lens system used in air is slightly less than unity; by using oil immersion objectives, N.A. can be made nearly equal to 1.5. Consequently the best we can do is to have

$$d = 1/3 \cdot \lambda_0$$

The wave length of the light near the center of the visible spectrum is 6×10^{-5} cm., and so the smallest possible distance, *d*, for visible observation is approximately 2×10^{-5} cm., *i.e.*, 1/50,000th cm., or 1/125,000th of an inch. This is then also the diameter of the smallest particle that we can see or photograph with the aid of the most powerful microscope, using ordinary visible light.

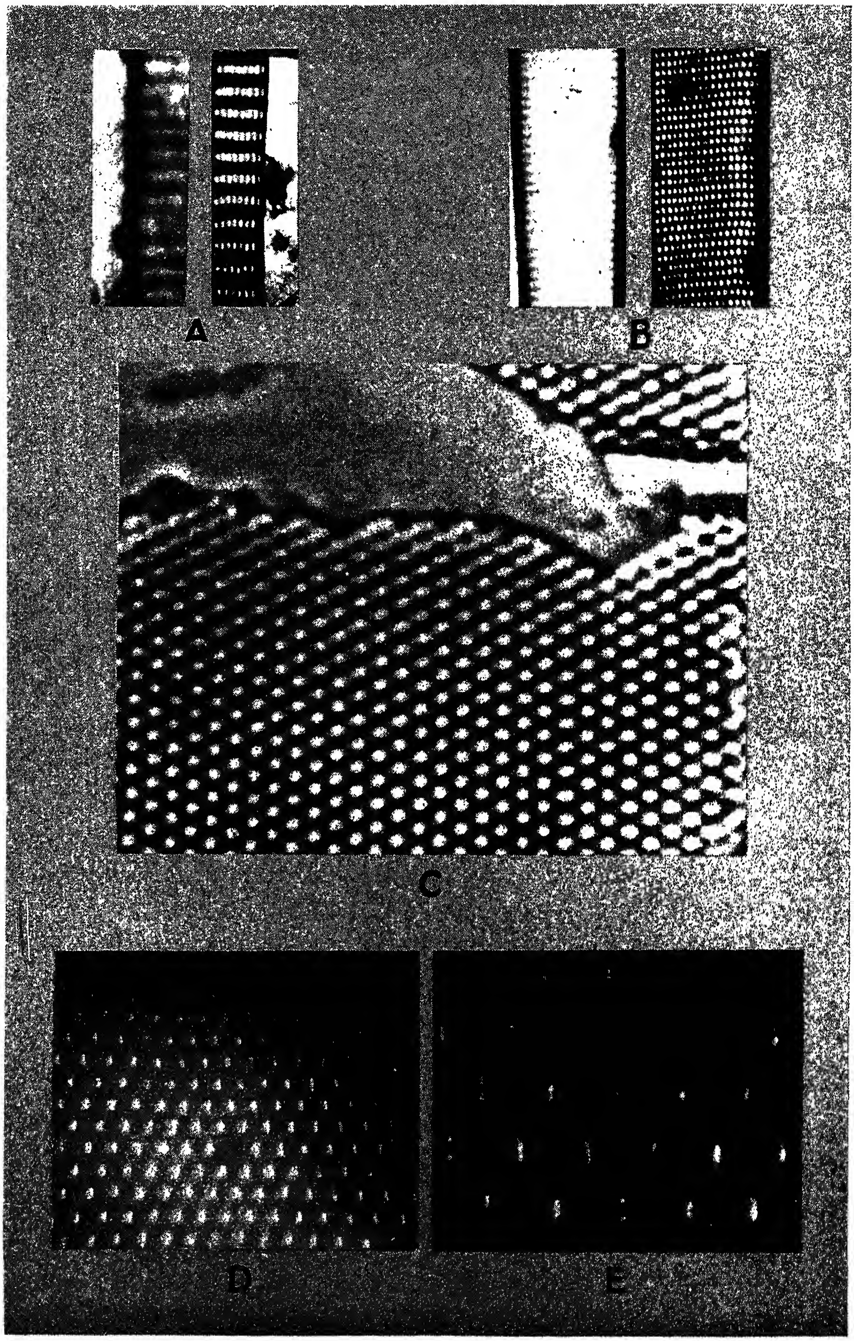
If we make use of ultraviolet light and photography, we can take pictures of particles about one-half this size and so have visible in the photograph greater detail in the object.

This marks the impasse reached by the microscopist by 1900.

Resolving Power and Magnification

The argument presented in the preceding section tells us that, with the help of even the finest microscope, the smallest particle that we can see must have linear dimensions of at least 1/125,000th of an inch. No amount of magnification or re-magnification can reduce this minimum value. Where then does magnification come in?

Since the naked eye cannot see differences in dimensions smaller than 1/250th of an inch, it is at once apparent that a


A
B
C
D
E

Explanation of plate on page 40

The purpose of these pictures is to show the manner in which the electron microscope contributes to the interpretation of the best optical microscopic pictures.

The objects are diatoms of a type often used for testing first class oil-immersion objectives.

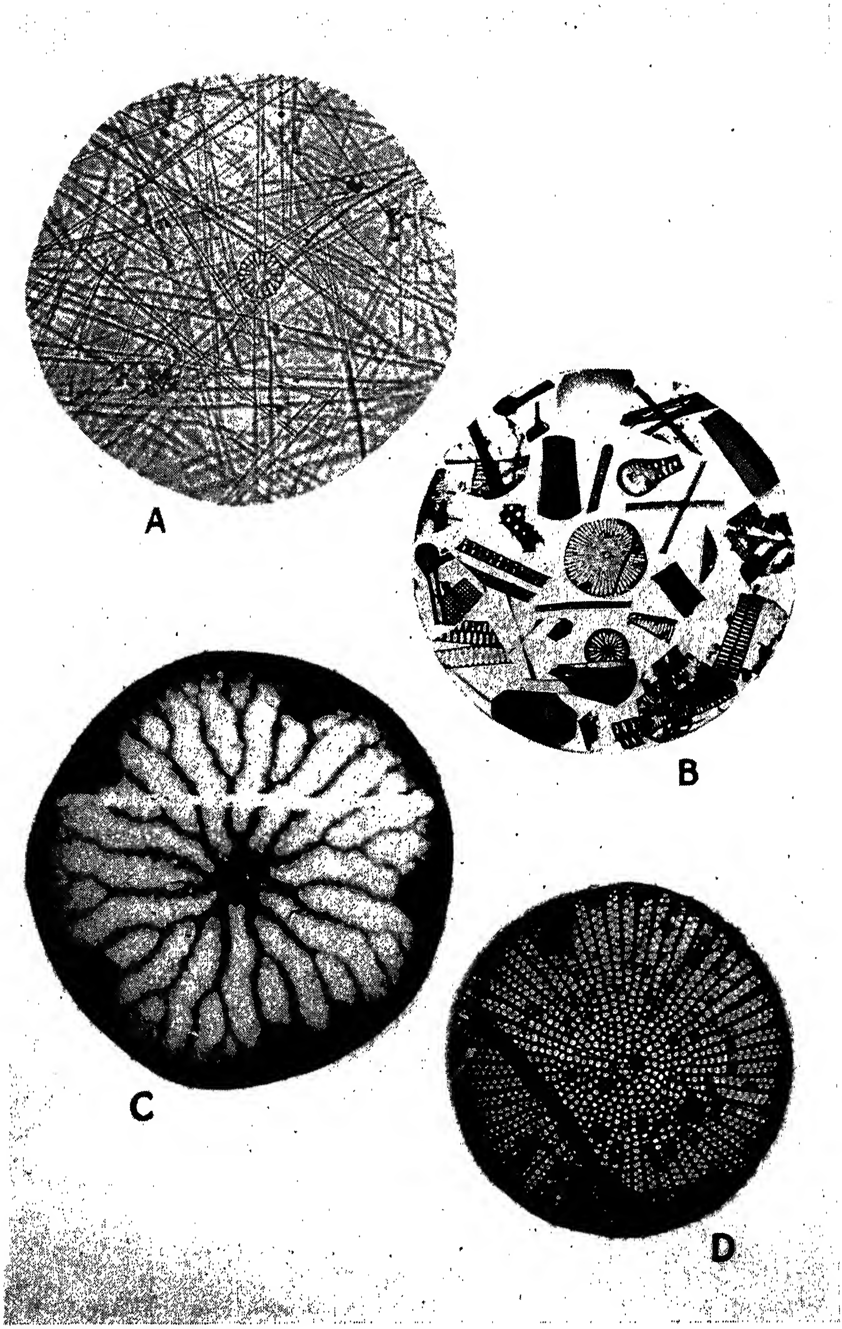
A shows photographs of *Synedra delicatissima.* For the oil-immersion pictures, perfectly symmetrical Köhler illumination and an aperture giving N.A. 1.4 were used. The optical microscope (left) indicates the presence of rows of holes, but the electron microscope (×5,000) (right) goes further than this and shows positively the size and arrangement of these rows and holes. The holes are approximately 0.5μ in diameter, they are separated in the row, center to center, by a distance of 0.20μ; the rows themselves are approximately 0.90μ apart.

B shows corresponding photographs of the diatom, *Amphipleura pellucida.* The optical pictures (left) were taken with the same Köhler illumination and the same N.A. 1.4. The optical microscope shows bars in the valves and hints at the presence of openings within the slots. The electron microscope (×5,000) (right) shows clearly the small holes, approximately 0.14μ in diameter, with centers separated by 0.20μ. The rows are about 0.20μ apart.

C shows a valve fragment of the well-known test diatom *Pleurosigma angulatum* photographed with symmetrical illumination and an aperture of N.A. 1.0. The magnification is enlarged optically to ×5,000.

D is an electron picture of a portion of **C** with magnification ×5,000. That **D** itself does not tell the whole story given by the electron microscope is shown by the fact that **E**, a portion of **D** with total magnification ×15,000, shows that each individual hole in the shell is really a series of four holes approximately 0.1μ in diameter separated from each other by 0.03μ. The holes appearing at lower magnification are about 0.7μ apart.

The optical pictures were made by Dr. D. H. Hamly of the Department of Botany, University of Toronto.



A B C D

Explanation of plate on page 42

This plate reproduces pictures of various kinds of diatoms.

A is an optical photograph (by Dr. L. T. Newman) of a selected specimen of the type *Synedra delicatissima* kindly supplied by Dr. Paul S. Conger of Washington, D. C. Note that there is an odd circular type of diatom quite foreign to the others. The magnification is ×2,000.

B is a made-up slide showing a random collection of electron microscope pictures of various types of diatoms found in a random sample of diatomaceous earth. The elongated ladder-like structures are similar to those types shown in A.

C is an electron microscope picture of the diatom type, *Cyclotella sp.;* magnification ×25,000.

D is an electron microscope picture of the diatom type, *Cyclotella sp.*, magnification, ×5,000.

dimension of 1/125,000th of an inch must be magnified at least 500 diameters. This is usually written $\times 500$. Though the eye can see very small detail, eye strain and fatigue are avoided by increasing this magnification from 2 to 6 times. This increase in magnification is dependent upon the initial contrast between object and field which, in the case of small transparent particles, is very low. When the contrast is low no increase in magnification will increase the amount of detail, since image contrast decreases with increase in magnification. Magnifications of $\times 1200$ to $\times 1800$ are practical with well-stained materials and a good oil-immersion objective, but higher magnifications up to about $\times 5000$ are possible only with metallurgical materials showing exceptional contrast.

Contrast is a primary requirement in the visual perception of any image. When contrast is low it can be increased optically by the use of suitable filters and photographically by the proper use of emulsions and developers. Enlargement reveals very small detail by bringing it to the size perceptible to the eye. Empty magnifications and low contrast result when the limits already mentioned are exceeded.

The detail seen in unstained materials, such as cleaned diatoms, can be improved by increasing the contrast with the use of media of higher refractive index than that of cedar oil. Where the use of ultraviolet light is possible with unstained organic materials, such as thin cross-sections of tissues and of bacteria, the peculiarities of differential fluorescence and absorption increase the contrast. While these special techniques have been very successful in revealing detail up to the limits of resolution, in ordinary practice these limits are very seldom attained.

The Ultramicroscope

About 1900 a completely new technique in microscopy was developed by Zsigmondy and Siedentopf of the Zeiss Company. Like many other scientific advances it involved the application in a new field of an ordinary well-known experimental fact.

We are quite well aware that the air of any ordinary room is filled with motes and dust particles which are invisible when the room is well lighted. However, if we darken the room as a whole, admit a narrow beam of sunlight or throw a narrow beam from a projection lantern and view this bright beam from the side, we are made aware of the presence of thousands of motes or dust particles floating in the air. But if we look in and along the beam toward the source, the same dust particles or motes are quite invisible.

The reason for this phenomenon is that the small motes or particles scatter or diffuse or reflect in all directions the light which falls on them. When we view the beam *laterally* our eyes catch the scattered light against a dark background, and so we say we *see* the particles. When the whole room is again illuminated the dark background disappears so that we can no longer distinguish the light scattered by the particles.

Fig. 25 shows diagrammatically the essential features of the Zsigmondy-Siedentopf ultramicroscope. The sample of material containing the small particles—gold ruby glass or some other

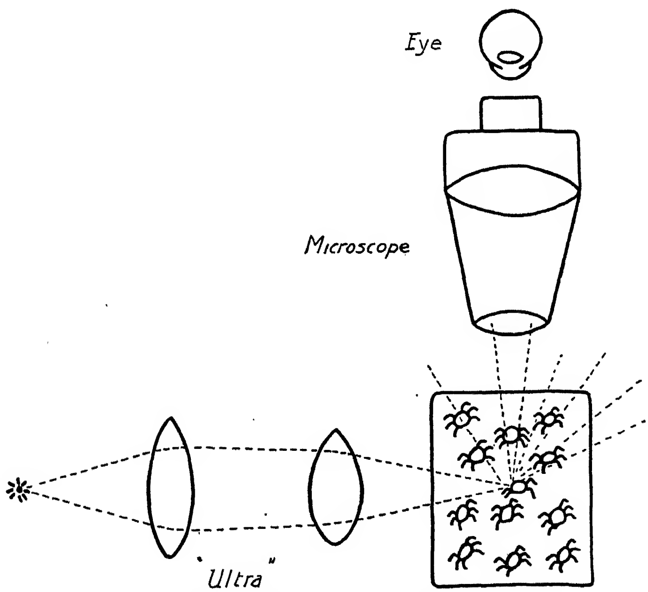


FIG. 25. The principle of the ultramicroscope.

colloidal solution—is illuminated in one direction and viewed at right angles to the illuminating beam by means of an ordinary microscope. The 'ultra' part of this instrument is merely the unique illumination which enables the particles to scatter light into the microscope so that they are seen against a dark background.

One deficiency of such a set-up is that we are only made aware of the presence of such particles, but we cannot measure them or tell their shape. On the other hand, we can be made aware of the presence of particles very much smaller than those set by the limits given in the previous sections. The reason for this is that particles of any size will scatter or diffuse some light in all directions, and whether we see such particles or not depends on whether the amount of light scattered from a particular particle, which manages to enter the microscope and consequently the eye of the observer, is sufficient to affect the retina. This is chiefly determined by the brightness of the illuminating beam. It has been calculated that we should be able to see ordinary molecules with the ultramicroscope if we could illuminate them so that they would be as bright as the surface of the sun.

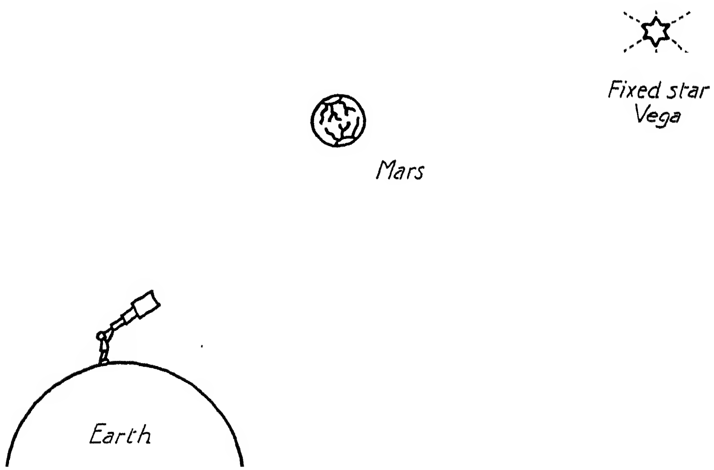


FIG. 26. Planets are so near that a telescope shows details of their surfaces, and gives a distinct measure of their sizes; but fixed stars appear only as points of light, on account of their enormous distances from the earth.

This phenomenon is analogous to one encountered in telescopic vision. The only heavenly bodies for which we can determine the size by direct observation or of which we can see any detail are those belonging to our own solar family, such as planets and their satellites (Fig. 26). But we are made aware of the existence and position of the far-away fixed stars because they are so brightly illuminated that our telescopes gather and transmit to our eyes sufficient light to affect the retina. We are however not able to tell anything about the size and shape of these fixed stars by vision alone, no matter how huge our telescopes.

Chapter 3

What is Light?

The Meaning of a Scientific Theory

In the preceding chapter we have given some answers to the question, "What does light do?" The present chapter will deal with the more difficult question, "What is Light?"

The answer to this question, as to any similar one in the realm of science, depends on the amount of information which has been accumulated from experimental facts. This information tells us what a certain thing does, but the nature and substance of the thing itself remains a mystery until the human mind applies itself to the interpretation of the experimental facts and forms a theory.

What do we mean by a scientific theory?

We have many natural phenomena to theorize about. What is heat? What is sound? What is light? What is gravitation? And so on. In each case we try to construct in our imagination a mechanical model which will, in its working, produce the observed effect.

What do we mean by a mechanical model? We play around with various machines and gadgets which do things for us. We know the forces involved in springs; we experience the pulls in a tug-of-war; we are familiar with the flow of liquids through pipes; the intricate motions of tennis balls, baseballs, and golf balls are matters of everyday experience. So when we come to account for such an intangible, nebulous thing as, for example, heat, which we say "flows" from one place to another in material bodies, we try to imagine *what* is flowing. The earliest theory or model of heat was the picture of an imponderable, invisible, elusive fluid, known by the name, *caloric*, which was supposed to flow from a hot body to a colder body. No experiment was ever devised

which could really make such a fluid evident to our senses. Of course, everyone now says, "Heat is a mode of motion," by which we mean that we think we have a more satisfactory mechanical model for heat; we say that "heat possessed by a piece of matter is just the sum total of the energy of motion of the molecules, and when heat is transferred from one place to another what is really happening is that fast-moving particles are communicating their motion to slower-moving molecules and making the latter move more quickly." In our first rough picture we imagine that these particles act like swarms of minute tennis balls, colliding and bouncing back in a random manner.

Our judgment as to what is a *true theory* is merely our answer to the question, "Which is the most satisfactory mechanical model?" or, "Which model, acting in agreement with the laws which our machines must obey, will best foretell for us experimental happenings?"

We cannot say that any scientific theory is *true* or *false*; we decide merely which mechanical model is experimentally the most satisfactory. This is the sense in which we shall speak of a *theory*.

Theories of Light in Ancient Times

During the earliest period of recorded human speculation about the nature of light very few optical facts were known. People saw with their eyes the things around them and a few inquisitive minds wondered how the process of vision came about. The early Greek philosophers attempted explanations of the phenomenon of light, but their experimental knowledge was confined to three points: (1) transmission of light in straight lines, (2) reflection from smooth polished surfaces, and (3) bending of light by refraction in passing from one medium to another.

Pythagoras (582-497 B.C.) taught that vision was caused by particles continually projected from the surfaces of visible objects into the pupil of the eye. The teaching of Plato (427–347 B.C.) was rather more elaborate; he held that vision was brought about by the union or interaction of something emitted from the eye itself and something else emanating from the object itself. Plato's

model was threefold; a stream of divine fire emitted by the eye itself became united with an influence coming from the sun and this combination reacted with a third emanation from the object itself and so accomplished vision. The simpler Pythagorean idea survives today in what is known as the *corpuscular theory* of light.

Aristotle (384-322 B.C.) discarded these emission theories of light and substituted for them what might be looked upon as a nebulous forecast of the modern *wave theory*; he maintained that light was merely a quality of, or action in, a space-filling medium which he called the *pellucid*—the forerunner of the modern ether.

Modern Theories: Corpuscular Theory and Wave Theory

Little would be gained by trying to unravel the intricacies of the theorizing regarding light down through the ages. The outburst of experimental science beginning in the sixteenth century revealed in the course of one century many of the fundamental properties of light and laid the foundations of all future work in this field.

In 1608 the first telescope of which there is a published account was constructed by a Belgian spectacle-maker, Hans Lippershey. One year later, using this as a starting point, Galileo built a telescope of such perfection that he was able to see, for the first time in human experience, the satellites of Jupiter; this opened a new field to the astronomer.

In 1621, Willebrod Snell, a Dutch scientist, discovered the laws governing the refraction of light, the phenomenon illustrated in Figs. 2 to 5. These laws enable us to foretell the amount of bending undergone by a light ray as it goes from one medium to another.

In 1675, Olaf Roemer, a Dane, discovered that light is propagated through space at a finite, though very large, velocity and not instantaneously, as previously supposed. He made this discovery by observing that the times of the eclipses of the satellites of Jupiter did not show the regularity that one would have

expected. He accounted for this irregularity by assuming that the differences were due to the fact that, on account of the changing distances between the earth and Jupiter, the light took a longer or shorter time to come from Jupiter to the earth as the distance between these two planets changed due to their individual rotations about the sun. He found that this anomaly was completely accounted for if one assumes that light travels through space at the rate of 192,000 miles per second—the first experimental determination of this important quantity (Fig. 27).

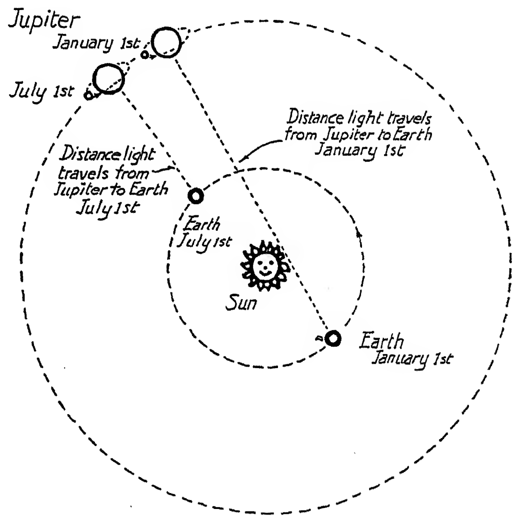


FIG. 27.

Roemer's determination of the velocity of light.

About the same time (1672) Sir Isaac Newton made the remarkable discovery that ordinary white light, such as sunlight or light from an incandescent solid, is broken up by a prism into many colored components, giving the colors of the so-called spectrum—a distribution of colors already familiar in the rainbow; in fact this discovery of Newton's at once offered an explanation of the formation of rainbows.

To complete the contributions of this wonderful century of progress we record that in 1690 Christian Huyghens, also a Hollander, observed that the natural crystal, Iceland spar, could break up a single beam of light sent through it into two parts, which

travelled through the crystal in different directions. This discovery revealed the phenomenon which we call polarization of light, a property of light which has become a matter of everyday knowledge since the invention of polaroid. The striking thing about this phenomenon is that if a beam of light is passed successively through a pair of such crystals the light corresponding to one or other of the two parts can be shut off by rotating one of the crystals about the direction of propagation of the light, although each of the crystals is of itself quite transparent independently of any such rotation (Fig. 28).

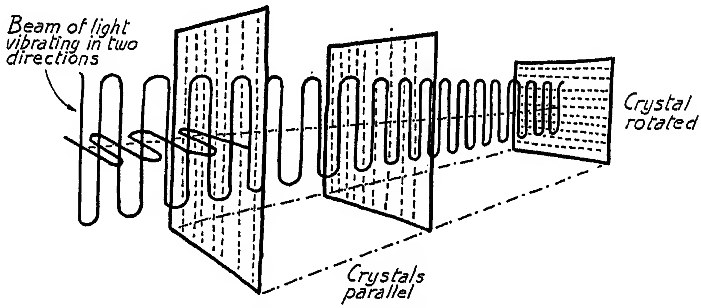


FIG. 28. Polarization of light by crystals in Polaroid.

The construction of a theory which would be able to explain satisfactorily all these phenomena was a much more difficult task than that which had confronted the Greek philosophers. It is small wonder then that many different theories, upheld by different schools of thought, co-existed for long periods of time.

In Huyghens' and Newton's time two opposing theories existed; these were known as the Huyghens wave theory and the Newton corpuscular theory. On close examination of Newton's own statements it will be found that his theory was the more general one and that it left room for both the corpuscular and the wave concept.

In so far as light is a transmission of energy through space we would be justified in accepting a *corpuscular theory*, that is, that a light source shoots out in all directions fine particles or corpuscles which enter our eyes and affect the retina (Fig. 29).

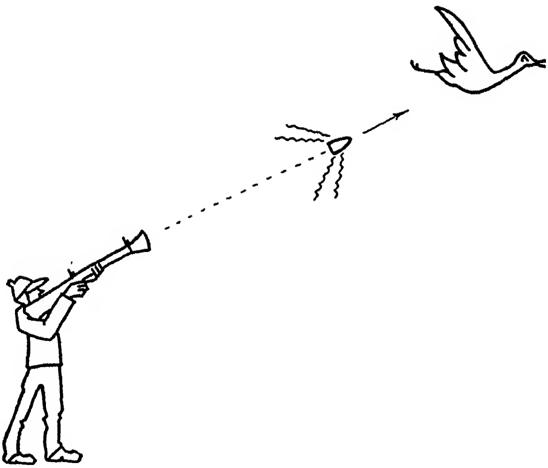


Fig. 29. Energy may be transmitted through space by a moving mass.

But there is a way in which energy can be transmitted through space without the actual transmission through space of any particles or corpuscles. One of the first experiments in physics that almost every child does illustrates this fact (Fig. 30). A row of dominoes, each standing on end and near its neighbor, can be arranged in any fantastic figure and, when one end domino is toppled over, a pulse of falling dominoes passes along the line. The motion, and therefore the energy, is carried along through space, but matter is not transferred from the starting point to

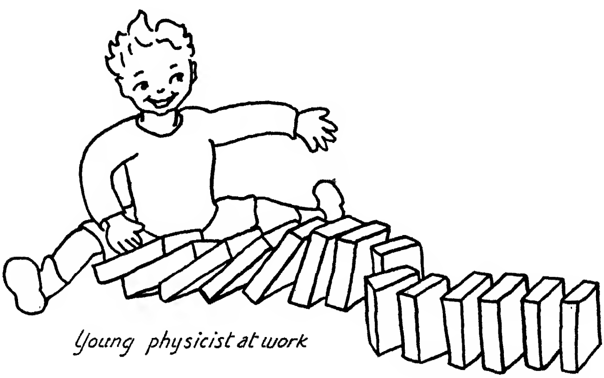


Fig. 30. The transmission of energy from point to point without the transfer of any mass.

the end point. Just as simple, but more far-reaching, phenomena are observed in the waves on the surface of water. These waves are caused by the winds disturbing the equilibrium of a quiet, flat surface, and we know from the destructiveness of violent storms that these waves carry along enormous amounts of energy. But if we toss a stick on the surface at any point we see that it is not carried along by the wave motion but merely bobs up and down close to the point where it was thrown in. From this we gather that the water itself is not carried along with the wave motion, and that consequently, energy alone must be transmitted.

We may say then that, as far as light is merely a transmission of energy, we would be justified in accepting a *wave theory*: that a light source initiates a vibrational motion in some medium, and that energy is carried along without the passage of material particles along the line of the motion.

Although we have outlined above the approaches to a solution we are still not very convinced that we can answer the question, "What is Light?"

Chapter 4

Wave Motion and Wave Motion Media

The Elements of Wave Motion

As indicated already in Fig. 23 (p. 36), our fundamental ideas regarding the propagation of energy by wave motion come from the actual motion which we observe in waves on the surface of water. Fig. 31 illustrates what happens when we drop a stone on the smooth surface of still water. As the stone strikes the surface, wavelets spread out in concentric circles from the point of contact as center, and a floating object on the surface at any point will bob up and down. Fig. 31 may be taken to represent graphically two facts about this wave motion. First, if we could

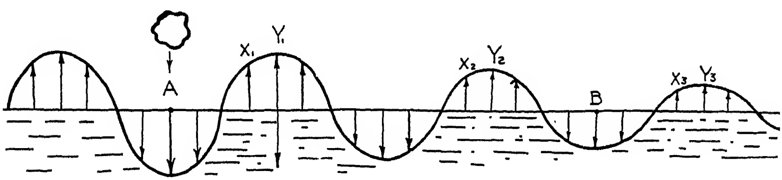


FIG. 31. Graph of water wave showing the meaning of wave length, phase, frequency and velocity for a typical wave motion.

imagine that the surface could be kept fixed as it is at any given instant, as when the moving picture director cries, "Hold it!," the curve shows the relative position of the surface points at that instant; in the second place, if we imagine the curve travelling along in the direction of, for example, A to B, with the velocity of the wave itself, we have a picture of the progress of the motion; and, if we fix our mind on any one point, as Y_1, we get a true idea of the motion of a water particle at this point as the wave sweeps along. Since the actual motion of the particle is at right

angles to the direction in which the wave itself is progressing, this is said to be a *transverse wave*.

We can learn from this figure some of the technical language applied to wave motion in general. As we have already pointed out, the distance from crest to crest or from trough to trough is called a *wave length;* two points which are in corresponding positions in the crests or troughs, as X_1 and X_2, are said to be in the same *phase*. The number of times the particle at a point on the surface bobs up and down in one second, or, as we might say, the number of times a particle completes a cycle in a second, is known as the *frequency* of the wave motion. It is at once apparent that the velocity with which the disturbance moves forward is equal to the number of wave lengths marked out per second, *i.e.*, is equal to the frequency multiplied by the wave length. This is expressed in mathematical shorthand by the symbols

$$V = \nu\lambda$$

where V means velocity, ν frequency and λ wave length.

Sound Wave Motion in Gases and Liquids

Now although we get our simplest example of wave motion from surface water waves, this phenomenon is quite inadequate to serve as a model of wave motions propagated through space in three dimensions. The water wave spreading over a surface is confined to that surface and cannot serve as a model for the motion caused by the explosion of a depth bomb below the surface, which spreads out from the source in all directions in the ocean, nor for the disturbance set up in the air by, say, a firecracker. These two are examples of wave motions in *three* dimensions; but in both cases we have energy being transmitted from one point to another in the form of sound.

For a more detailed description of the phenomenon we illustrate, in Fig. 32 (a, b, c), the propagation of the sound from a gong which has been struck by a drum-stick; this drawing was constructed in accordance with our accepted model of the air, namely, that it is a collection of small spherical particles swarming

around in random motion. The metal disk vibrates when struck. As it moves to the right it compresses the air in its immediate vicinity to the right; then it swings back toward the left, making a compression to the left side and leaving a rarefaction on the right side. The air near the disk goes through this succession of oscillations as long as the vibrations of the disk continue.

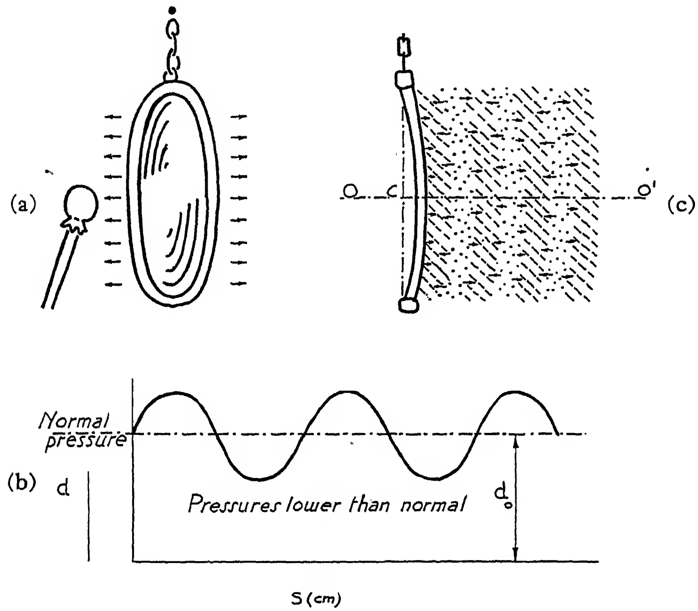


FIG. 32. The motion of air particles in the neighborhood of a vibrating gong. As the sound travels to the right, (b) represents how the pressure varies at any point along CO′.

The events taking place on the right side of the disk are shown diagrammatically in Fig. 32c which represents, as it were, a snapshot exposure of the air particles. The shaded regions are those of compression and the dotted regions are those of rarefaction; the little arrows indicate the direction in which the air particles have moved just before the snapshot was taken. A moment later the disk will bulge to the left and the state of affairs will be reversed in every respect. We can really measure the pressures along the line CO′ as the sound is travelling, and the

curve in Fig. 32b shows graphically the relation between the pressure at a point along CO′ and the distance of the point from the center of the disk at the time of the snapshot. As the direction of motion of the air particle is the same as that in which the sound moves, this form of wave motion is said to be *longitudinal.*

The human ear, which acts as a receiver, responds to the variations in air pressure, and the sensations produced physiologically are interpreted as sounds of different loudness and pitch. Since sound cannot be transmitted in a vacuum, we must accept the conclusion that the air is the mechanical link between the sound transmitter and receiver.

That sound is a form of energy is readily understood if one recalls that windows may be broken by the impact of sound waves resulting from an explosion at some distance. Although the results are not so violent, energy is also necessary to produce the motion of our ear drums as they record sounds falling upon them.

Sound travels not only through air (and other gases) but also through liquids and solids. In every case the origin of the sound is a periodic disturbance which causes a slight displacement of the molecules of a material medium, and the resulting periodic wave motion is due to the action of forces which always come into play to push these disturbed molecules back into their original positions. The velocity with which the disturbance moves through any medium depends on the nature of the medium, and the molecular forces called into play to restore the molecules to their original positions are known as *elastic forces.* We call this property of a medium, elasticity; every medium has its own peculiar elastic properties which are indicated by so-called *elastic constants.*

Sound waves in any medium are thus elastic waves. If the frequency of the motion set up in any medium is between 20 cycles and 20,000 cycles per second, the resulting motion affects our ears, causing a sensation in the brain which we call sound. Any disturbance set up in a given medium, whether we can hear it as a sound or not, travels with the same kind of vibratory motion and with the same velocity. In fact, the velocity of any such disturbance in any given medium depends only on the elastic con-

stant of that medium and its density, and not at all on the sound characteristics, such as pitch and intensity.

Consequently, a disturbance caused by an explosion in air travels with the same velocity in still air as the sweetest note of a bird. Again, when a depth charge is set off in the ocean at a considerable distance from the surface, the disturbance travels out from the center of the explosion in spherical waves of compression and rarefaction with a velocity which is the same as the velocity of any ordinary sound in water, or about 4,500 feet per second. The direction of the to and fro motion of the molecules in liquids is always back and forth along the direction of progress of the disturbance, so that here, as in the case of gases, we have *longitudinal* wave motion. This is the only possible kind of wave motion that can exist in liquids and gases.

Sound Wave Motion in Solids

When we come to study how sound travels in solids, we find the complication that two kinds of sound waves are possible. If we pluck a violin, harp, or piano string in the ordinary way we obtain a rich, pleasing note: but if, instead of plucking the string, we pull it with a resined cloth along its length, we obtain a screeching, unpleasant note. Both these notes are due to vibrations set up among the molecules of the string; but from the manner in which the disturbance is caused we see that the motion of the molecules, in the first case, must be perpendicular to the length of the string or *transverse*, whereas in the second case the motion will be *longitudinal*, *i.e.*, parallel to the length of the string. So if we could see the motion of the molecules inside the string, we should observe that the disturbance is carried along in one case by a *transverse wave* and in the other by a *longitudinal wave*. Experiment has shown that these two wave motions in solids are regulated by two separate elastic constants, which correspond to two separate elastic properties of the solid of which the string is made.

In summarizing our various experiments demonstrating the propagation of different waves in material media we should search

for a common criterion that makes the establishment of the waves possible in all cases. To begin with, the equilibrium of the medium in question is in every case disturbed by an external force which imparts to the particles of the medium a displacement in the direction of the force. This displacement is not permanent. The medium offers a resistance which tends to restore the equilibrium position of the particles when the external force is removed. Plastic solids, such as plasticine and dough, which do not possess these strong molecular forces, cannot maintain a wave motion and consequently cannot transmit sound. The motion of the diaphragm of the gong in Fig. 32 sets up longitudinal waves in the air due to the ability of the gong to vibrate for some time after being hit by the drum-stick. In this case we have not only the vibrations in the air but also a wave motion travelling to and fro within the metal of the plate at a definite rate—a wave in the solid itself.

It is to these internal wave motions that we referred in pointing out the existence of wave motions in the material of the violin and piano strings. The disturbances which we call earthquakes are wave motions transmitted through the solid crust of the earth, and it is this kind of wave motion in a large volume of homogeneous solid that has really served as the prototype for the wave theory of light.

The Velocity of a Wave Motion

The interrelation between the elastic constants and the density of solids and the velocity of a wave motion was very fully explored many years ago, beginning with the experiments of Robert Hooke (1635-1703). It was found that the velocity, V, was given by the relation

$$V = \sqrt{E/d}$$

where E and d are the elastic constant and density of the solid, respectively.

In Table 2 is collected some information regarding the velocity of sound in some typical media, as deduced from this relation; in every case the numbers obtained for the velocities agree very well

with those obtained from experiment. Of course all these quantities change with changing temperature; usually these measurements are taken at a standard room temperature, 20° centigrade or 68° Fahrenheit.

Table 2. Values of the Velocity of Sound.

Substance	Elastic Constant (dynes per sq. cm.)	Density (gram per cc.)	Velocity of Sound * (cm. per sec.)	(ft. per sec.)
Aluminum	7.0×10^{11}	2.70	510,400	16,740
Copper	12.3×10^{11}	8.93	356,000	11,670
Lead	1.6×10^{11}	11.37	122,700	4,026
Steel	20.9×10^{11}	7.8	499,000	16,360

* From Smithsonian Tables.

Courtesy RCA Laboratories

Trachea of honey bee (×15,000).

Courtesy RCA Laboratories

Trachea of mosquito larva (×7,500).

Chapter 5

The Wave Theory of Light Accepted

The Medium Necessary for Light Wave Motion

Since the conception of a wave theory of light was based on experimental proofs of the existence of wave motions in material media, it was incumbent on the supporters of this idea to provide a model medium in which such a wave motion could take place. It was postulated that a satisfactory medium did exist throughout all space; this was given the name *ether*—just another of the invisible, imponderable fluids which have often served to satisfy the human longing for an explanation of mysterious natural phenomena.

Manifestly this new medium must have an elastic constant and a density such that the velocity of light would be given by the relation

$$V = \sqrt{E/d}$$

When first suggested, the ether was assumed to be a fluid, *i.e.*, a liquid or a gas, and so the motion of the particles of the ether was supposed to be longitudinal.

Modern determinations of the velocity of light make it 186,000 miles per second (or 30,000,000,000 cms. per sec.)—a perfectly enormous velocity compared with any that had ever been observed for any sound wave. In order that such a velocity should exist in any medium the value of the elastic constant, E, would have to be extremely large or the density of the medium would have to be extremely small, or both of these conditions would have to obtain at the same time to a lesser degree. In effect, if the ether were supposed to have an elastic constant, E, equal to that of steel, *i.e.*, if it were as rigid as steel, the density could not be larger than

about 1/50th that of the lightest known substance, hydrogen gas; if, on the other hand, the ether were supposed to have a density equal to that of hydrogen, the elastic rigidity would have to be about 50 times that of steel, one of the stiffest materials.

The Corpuscular and Wave Theory Contrasted

We have seen that in so far as light is to be considered merely as an example of the transmission of energy through space, we have no particular reason to prefer one of the theories to the other. Table 3 offers a comparison of the two theories and an appraisal of the degree with which they meet the demands of experimental facts.

Table 3. Comparison of the Two Theories of Light.

Phenomenon	Experimental Facts	Wave Theory	Corpuscular Theory
High velocity	186,000 mls./sec.	Medium necessary but very difficult to conceive	Satisfactory Medium not necessary
Propagation	In straight lines	Difficult	Satisfactory
Reflection	Experimental laws	Satisfactory	Satisfactory
Refraction	Experimental laws	Satisfactory	Questionable
Diffraction	Bending around obstacles	Satisfactory	Impossible
Interference	Possible for two portions to be added to produce darkness	Satisfactory	Impossible

We realize from this table that neither theory is able to give a satisfactory account of all experimental facts. If we accept the existence of an ether having very freakish properties, the wave theory seems to afford a better model than the corpuscular theory. The latter fails completely to offer any explanation for diffraction or interference. As we have already dealt with diffraction (Figs. 19 and 20, pages 33 and 34), we shall now consider interference in more detail.

Interference of Light Waves

The simplest experiment showing interference of light is illustrated in Fig. 33. Light from a source, S, is shielded by a screen with a small hole, P, which acts as a point source for the space to the right of the screen. The light from P spreads out and falls on a second screen which is pierced with two holes, A and B, symmetrically placed with respect to the center line of the figure. The light passing through the holes, A and B, is allowed to fall on a third screen some distance away, so that the cones of light from A and B will overlap. Now, instead of the last screen being uniformly illuminated, we observe on the screen alternate bands of light and darkness. That is, we have the curious phe-

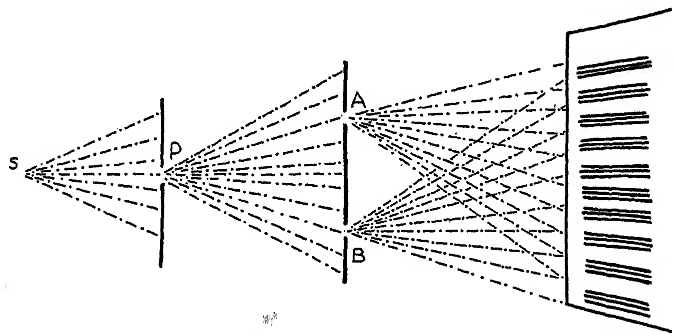


FIG. 33. How two light beams "interfere" with each other.

nomenon of light from two sources, A and B, travelling independently to a screen and, with no obstacle involved, producing darkness at some points on the screen. By no stretch of the imagination can we convince ourselves that corpuscles travelling along can produce this effect.

On the other hand, the wave theory gives a very natural forecast of such a result. If one wave crest tends to produce a compression of the ether and the trough of a second wave tends to produce a rarefaction at the same point simultaneously, the two forces, if equal in intensity, will cancel each other and the ether will remain undisturbed at that point. If this happens all along the path of two co-existing waves, which thus we describe as opposite in phase, then the ether will remain undisturbed all along the wave

path. The effect is the same as if no waves had been propagated at all, or in other words we have two independent waves of light adding up to darkness. Fig. 34 illustrates this state of affairs by diagrams which correspond to those in Fig. 32b.

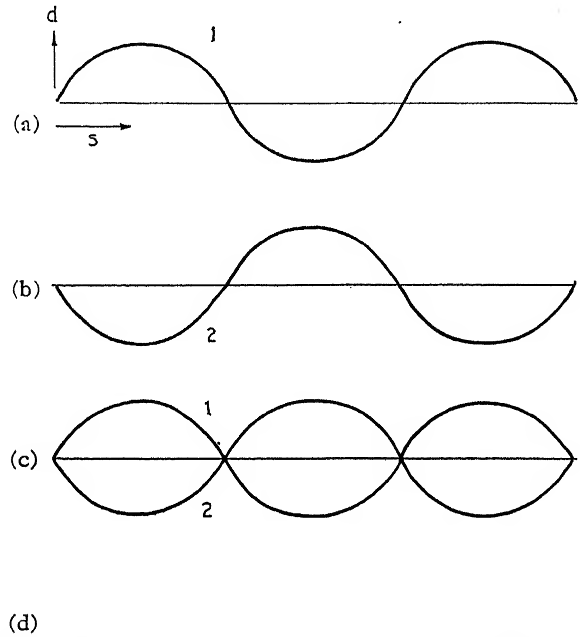


FIG. 34. The interference of two wave motions having a difference in phase of one-half a wave length.

These figures are snapshot exposures of the density of the medium along the direction, *s*, in which the light is travelling. Wave No. 1 starts with a compression, wave No. 2 with a rarefaction. It is apparent that one wave is a half wave length behind the other, or as we say technically, it is out of phase with the other by that amount. This leads to the complete cancellation of the two; if the two waves are superimposed as in Fig. 34c, the next figure, 34d, gives the graphical result, *i.e.*, no action at any point along the path. As each wave represents an independent light beam of the same intensity in both cases, we must conclude that light plus light adds up to darkness under the particular circumstances that were just described. If, on the other hand, we arrange it so that wave No. 2 in Fig. 34b is shifted with respect to the first wave by

a quarter of a wave length, we produce the conditions illustrated by Fig. 35 (a and b). Both waves have again the same wave length; they are shown superposed in Fig. 35c. When we combine these two so as to represent the distribution of compressions

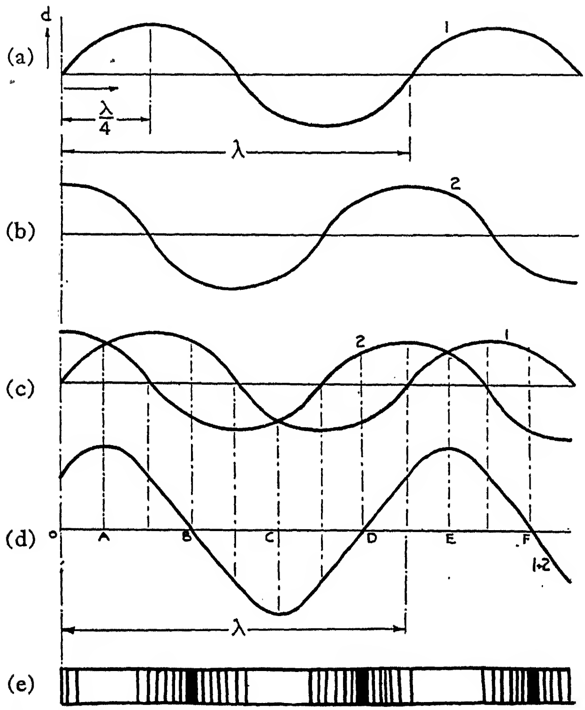


FIG. 35. The interference of two wave motions having a difference in phase of one-quarter of a wave length.

and rarefactions which is the resultant of the two, we obtain Fig. 35d, which is a new wave of different shape but still of the same wave length. Comparing Figs. 35d and 34d we realize that now the two waves do not cancel out all along the path from O to F as before, but only at the points B, D, and F. Between these points the medium is definitely disturbed and energy is propagated. If we were to expose instantaneously a strip of film along OF, this film, when developed, would show the pattern of Fig. 35e, *i.e.*, dark regions at B, D, and F, and light regions

between these points with shaded transitions. This example explains how nicely the wave theory accounts for the phenomenon of interference without any artificial assumptions.

The Acceptance of the Wave Theory

Although the attempt to explain the fact that light appears to travel in straight lines by using the wave motion model has always been a stumbling block, and still seems a bit artificial, we may say that by 1800 the wave theory was considered to have solved so many difficulties more acceptably than the corpuscular theory, especially as regards reflection, refraction, diffraction, interference, and polarization (somewhat later), that it was almost unanimously accepted as the *true theory of light*. But the medium, ether, with its unique properties, remained an enigma.

Up to a little after 1800 it was tacitly assumed that the wave motion involved in light was a longitudinal vibration, which means that the hypothetical ether had the properties of a fluid. But in order to explain the polarization of light (see Fig. 28), it was found necessary to consider the vibration as *transverse*, and, consequently, as only solids can have a transverse vibrational motion, scientists were constrained to look upon the ether as a solid; so during the last century there was a good deal written about the elastic solid theory of light. Of course it must have seemed a rash thing to think of such a solid pervading all space and still allowing the planets of the solar system to keep on moving through space about the sun without any sign of the slightest resistance being offered to their motion.

To conceive what at the time is looked upon as impossible is one of the most difficult steps that the human mind can take. We do not speak here of a sheer flight of fancy that scans the horizon of ideas recklessly, but of the matured thought which tears itself loose from seemingly unbreakable bonds and catches a glimpse of a new reality. In the realm of science these bold forward steps are the mark of genius. Such an advance is dictated by the accumulated knowledge of experimental facts, but at the same time these steps intuitively embrace new facts as yet undisclosed.

Chapter 6

The Electromagnetic Theory of Light

Faraday's Conception of Fields of Force

Up to the present we have been speaking of light as an isolated phenomenon of nature. It is probably not too much to say that, as far as physical sciences are concerned, the most important discovery of the nineteenth century was that the phenomena of light, electricity and magnetism are intimately related. In fact it has been proved beyond the shadow of a doubt that light is itself an electromagnetic phenomenon. The twentieth century has already gone a great way further in showing that the relation between light and matter is so intimate that it is possible to bring about a transformation of light (radiation) into matter and of matter into light. Indeed, both light and matter present a dualistic aspect—a wave motion and a corpuscular discreteness. In developing this idea we shall have first to recall the discoveries in fields of natural phenomena apart from but paralleling developments regarding light.

During the two hundred years preceding 1800 the fundamental facts regarding electricity and magnetism were known, in so far as the existence of two kinds of electricity and two kinds of magnetic poles is concerned. At this time it was not known that electrical charges and magnets had the slightest connection. Following a suggestion by Benjamin Franklin, the charge assumed by ebonite when it is rubbed with fur was called *negative* electricity, and that assumed by glass when it is rubbed with silk was called *positive* electricity. When two pith balls, one touched to charged ebonite and the other touched to charged glass, were brought close together they attracted each other; but if the two

were touched to the same electrically charged body (ebonite or glass) they repelled each other (Fig. 36).

Since there are forces existing between these charged pith balls, energy has been transmitted through space from one body to the other. The question arises, "How does this interaction of forces take place?"

This problem will naturally bring to mind the similar one involved in the propagation of light. There is, however, a difference between the two types of energy transfer. In the case of light, it is only a transfer of energy from a source to the receiver, which may be the human eye or a photoelectric cell or a thermocouple. It is a unidirectional transfer of energy, a flow in one direction, however it comes about. This state of affairs does not prevail in the case of the two pith balls.

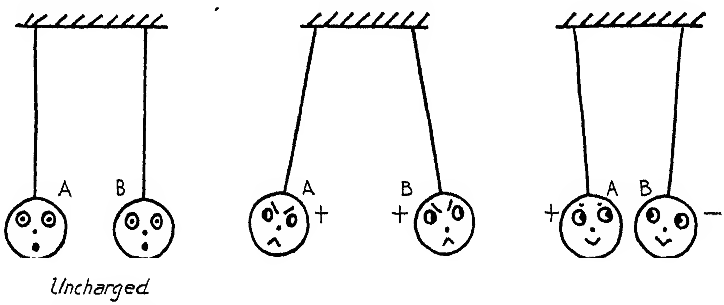


Fig. 36. How charged pith balls attract and repel each other.

The repulsion or attraction in the case of the two balls involves a certain amount of work, so that energy must have been supplied. But we cannot say that ball A is the source and B the receiver of energy or *vice versa*. Again, unlike light, this phenomenon does not involve a continuous transfer of energy from source to receiver, but the setting up of a state of equilibrium in which one pith ball is kept at a certain distance from the other in spite of the effect of gravity, which would make the balls hang vertically.

Most of the early workers were content to cover up their ignorance of this mysterious state of affairs by saying that it was just a case of "action at a distance," quite analogous to the mutual attraction of the sun and the earth.

An exactly similar situation developed with regard to the action of one magnet on another. In the minds of the early scientists, like poles repel one another and unlike poles attract one another due to some innate property of the magnets, not at all dependent on what is in the space between them.

Faraday, whose scientific thinking was dominated by mental pictures of mechanical models of natural phenomena, was constitutionally a vigorous opponent of the ideas involved in the phrase "action at a distance." He was a firm believer that the interaction between charged pith balls, as well as that between magnets, could be explained only by assuming that something was happening in the space surrounding the pith balls (or around the magnets); Faraday held that there was set up in the space a state of strain in some medium filling the space. The space around the reacting bodies thus became the mechanical link causing the repulsion or attraction. This state of affairs is expressed by saying that there is a *field of force* (electrical) around the charged pith balls and another different *field of force* (magnetic) surrounding the magnets.

Media for Electrical and Magnetic Fields of Force

These two media, electrical and magnetic, would have to be imagined to exist coincident with the ether which was assumed for explaining the propagation of light.

The whole question was complicated by the discovery of Oersted in 1820 that there existed a magnetic field in the space about a circuit bearing an electric current, and the discovery made later by both Faraday and Henry that currents of electricity could be made to flow in a circuit when the circuit was moved relatively to a magnetic field. According to the Faraday view all these effects will involve these electrical and magnetic media, and the idea grew that both electrical and magnetic effects might be considered to be due to the action of *one* medium, a kind of an electromagnetic ether.

Now to set up strains in a medium in space requires that the medium should have elastic properties, just the same kind of properties as the ether imagined for light. Faraday was brought to the

point of considering that *one* ether might suffice for all these purposes and suggested the hypothesis that *light itself was an electromagnetic phenomenon.*

Experimental Proof of Connection between Light, Electricity and Magnetism

Under the dominance of this idea, Faraday began to devise experiments which would establish the truth of this supposition. He searched for this evidence for many years and was rewarded for his endurance in the end by the discovery of what is known as the Faraday effect (1845). A beam of polarized light passing through some solids or liquids is affected by a magnetic field in such a way that the plane of polarization of the beam is rotated by a certain amount if the medium is subjected to a magnetic field. Faraday's search for the discovery of a similar influence of an electrostatic field upon light failed, but another British physicist, Kerr, in 1875, discovered that a solid or liquid becomes double refracting, like Iceland spar, under the influence of an electrostatic field. This is known as the Kerr effect; the Kerr cell, based upon it, has for many years been an important element in television.

Maxwell's Electromagnetic Theory of Light

Faraday's ideas found their most concise expression in mathematical terms in Clerk Maxwell's electromagnetic theory of light (1865). It achieves the fusion of the optical, electrical and magnetic theories, presented earlier, into one new theory that comprises all these phenomena. Faraday's ether, the carrier of electromagnetic energy and Fresnel's ether, the carrier of light, become one and the same; but the new ether is no more an elastic solid ether, but more etherial in nature. Maxwell maintains that light is propagated as a transverse wave in the ether. The existence of a wave demands that some physical quantity varys periodically along the direction of propagation of the energy transferred. This quantity, which formerly was the density of a material ether, now is the electric and magnetic field in free space. To speak of light as a transverse electromagnetic wave, then, means that along

the beam electric and magnetic fields exist which are directed at right angles to the direction of propagation and which periodically rise and fall in intensity from zero to a maximum to zero and a minimum and to zero again. Since Faraday's fields are made the variable quantity of the light wave, Maxwell's theory demands that electric and magnetic disturbances of the ether will show the same properties as light and will be propagated with the same velocity, i.e., with the speed of light.

This theoretical conclusion was verified experimentally by Heinrich Hertz in 1887. He found that electromagnetic waves were produced by a spark gap and could be reflected, refracted, diffracted, and focussed in the same manner as light waves and that they were indeed propagated with the same velocity. These epoch-making experiments thus confirmed Maxwell's electromagnetic theory in the most beautiful manner and everybody was very happy indeed. The mystery of light seemed to have been solved once and for all.

Chapter 7
The Electron

The Discovery of the Electron

In spite of the success of the electromagnetic theory, there were minor uncertainties to be cleared up. It was not clear, for instance, how dispersion of light in a refracting body came about. In order to develop a satisfactory explanation for these phenomena, L. Lorenz (1880) postulated the existence of small, electrically charged particles in all material bodies, which were to have the property of vibrating with a natural frequency when disturbed by a light wave. The creation of light at the source, *i.e.*, the excitation of the ether, also had to be ascribed to such vibrating charged particles. They were, as yet, of an entirely fictitious nature from an experimental point of view, but their existence was demanded by a brilliantly conceived and well founded theory.

It was not long before the physical reality of these charged particles was discovered. Early in the second half of the nineteenth century several outstanding physicists in various countries devoted a good deal of their time to the study of gaseous discharges, *i.e.*, the passage of electricity through a partially evacuated glass tube. Tubes known as Geissler tubes on the continent and Crookes tubes in England were the earliest examples of these experiments, and our neon signs are an end product of this chain. During this period we meet men like Johann Wilhelm Hittorf, Sir William Crookes, Eugene Goldstein, Sir J. J. Thomson, and F. Braun. Their work led to the discovery of the electron by Sir J. J. Thomson in 1897. All the workers which we have just mentioned had found that a certain radiation is given off by the cathode, the negative pole, when a voltage is applied through the

partially evacuated tube by connecting its sealed-in terminal wires to a Wimshurst machine, a battery of galvanic cells or, later, an induction coil. This radiation came to be known as the cathode stream or cathode rays.

Fig. 37 shows the diagram of a Geissler tube and the striations visible at a pressure of about 7.6 mm. of mercury.* Fig. 38 shows

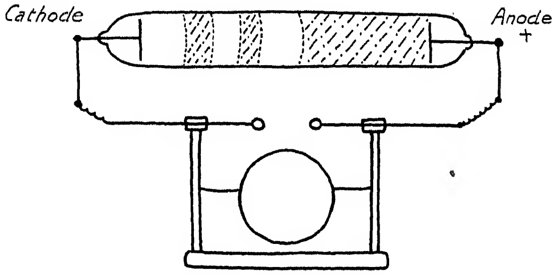


FIG. 37. Simple gas discharge tube.

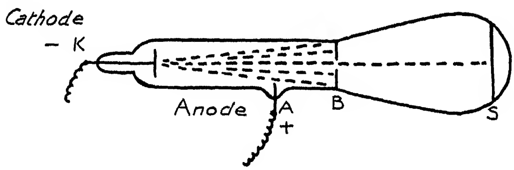


FIG. 38. The Braun discharge tube.

the familiar Braun tube, a sample of which is in almost any physics laboratory. It is generally about one and a half feet long. At K an aluminum disk is sealed into the neck of the tube to serve as cathode and at A a wire is sealed into the side of the neck to serve as anode. A disc with aperture at B permits passage of the central section of the cathode stream which emanates from the cathode, K. This fine beam then strikes a mica disc, S, which is covered with some luminescent powder that gives off light when struck by the cathode stream. The pressure in the Braun tube, first constructed by F. Braun in 1897, is about 5 millionths of one atmosphere or between 1/100 and 1/1000 mm. Hg. We

* Low pressure in evacuated tubes is expressed in terms of "Millimeters Mercury" (mm. Hg.). The normal pressure of the atmosphere at sea level is equal to 760 mm. Hg. 7.6 mm. Hg. is then about $\frac{1}{100}$ atmosphere. Modern radio tubes are exhausted to about $\frac{1}{100,000}$ mm. Hg.

may gather from Fig. 39 that the position of the anode is unimportant; it need not face the cathode. The linear emission of the stream from the cathode is not affected.

It was in 1895 that Roentgen made the discovery of x-rays while working with these gaseous discharge tubes, and it was a very live question at the time whether these two mysterious streams—the cathode stream and the x-ray stream—were some kind of light or not.

There were many ways in which the cathode stream resembled ordinary light. When the stream was allowed to fall on the surface of various minerals placed inside the vacuum tube, the minerals would be made to fluoresce, often with brilliant colors. Sir Wm. Crookes observed the phenomenon on such minerals as zincblende, fluorspar, willemite and diamond. When the highest vacuum was reached, the glass of the tube itself gave a brilliant fluorescence wherever the cathode stream struck it; the color was usually of a greenish hue, but the color was found to depend on the nature of the glass.

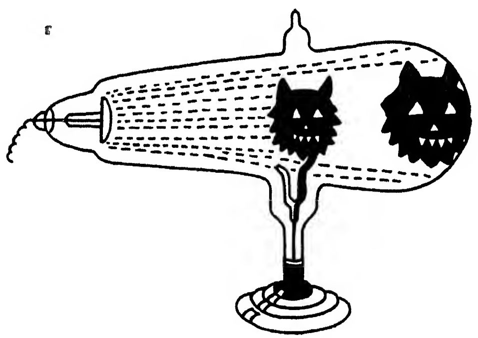


FIG. 39.

The shadow of a metal obstacle cast by cathode stream.

If an opaque object were placed in the path of the stream a very sharp shadow was cast (Fig. 39). This experiment suggests that the cathode stream travels in straight lines as in the case of ordinary light.

But there were two very important respects in which the cathode stream differed definitely from light. The stream was repelled from a negatively charged plate and attracted toward a

positively charged plate—which suggested that in some way the stream bore a negative charge. Again, if an ordinary small bar magnet were brought near the discharge, the stream was deflected at right angles to the direction in which the magnet was pointed. If the north pole of the magnet were brought near the stream, the deflection would be in one direction; for the south pole, the deflection would be in the opposite direction. These experiments are easily shown with the Braun tube (Fig. 38) as the motion of the beam is indicated distinctly by the movement of the fluorescent spot on the screen. No experiment was known which would show that ordinary light would be acted upon by either an electrostatic field or a magnetic field in such a simple manner.

The last experiment was epoch-making, because from it Sir J. J. Thomson deduced that the deflection was such as would happen if the cathode stream really consisted of a stream of *negatively charged particles*. By a combination of the application of electrostatic and magnetic deflection, Sir J. J. Thomson was able to determine the mass, charge and velocity of these particles—henceforth to be called *electrons*.

In the first experiments it was only the ratio, (charge on electron): (mass of electron), which was found, but neither the charge nor the mass could be found separately. Faraday's early work with electrolysis seemed to indicate that there existed an ultimate small unit of electrical charge, in magnitude equal to that on a hydrogen ion, the value of which was well known by 1897. Putting this value in for the charge gave a value to the mass of about 1/2000th of the mass of an atom of hydrogen, which was then our smallest known discrete mass. To suggest that a particle of such a mass really could exist was a most radical scientific departure, but Sir J. J. Thomson did not hesitate to take the plunge. And so was announced the discovery of the *electron*. Since, as we shall see, electrons of the same mass and charge may be produced from any kind of matter, we look upon the electron as one of the fundamental *building stones* of which all matter is made.

The Properties of Electrons

In addition to the properties already ascribed to the cathode stream, such as producing fluorescence, casting shadows and therefore travelling in straight lines, there are other properties of an electron stream which are of great interest.

We believe that we do not see the electrons or the electron stream; what was seen and was given the name of *cathode stream* was a visible *path* which the electrons produce in knocking against air molecules. The end of the electron beam is indicated by the fluorescence on a luminescent screen placed in the tube. It is by the use of these screens that visual observations of the motion of electron beams can be made. The modern television picture tube, an offspring of the Braun tube, utilizes this effect. Instead of natural minerals, the luminescent powders or phosphors, as they are called in this particular field, are produced synthetically by the most carefully controlled processes.

When electrons impinge on a metal foil, they give up their kinetic energy and produce heat. The foil may thus be brought to incandescence, a fact first demonstrated by Crookes. In our modern radio transmitting tubes the anodes must be water- or air-cooled in order to prevent excessive heating of the metal.

When electrons impinge upon a photogaphic plate they set free the silver from the silver bromide film in the same way as light does. The amount of blackening thus produced by electrons on a film indicates the intensity of the beam. This effect is utilized in high-energy cathode ray tubes in order to obtain a permanent record of very fast-moving beams. The pictures produced by the electron microscope are obtained in the same manner. It is evident that the photographic plate must be inserted into the vacuum chamber where the electrons move in order to be affected.

The Thermal Emission of Electrons by Metals

Electrons are put to so many uses in present-day electronic devices that it is a matter of primary interest to know how they are set free.

It has been mentioned above that all matter contains electrons. Among the solid substances, metals in particular show an abundance of electrons, and it is because of this fact that metals "conduct electricity." This latter statement simply implies that some of the electrons contained in the metal are free to move under the influence of an electric field established within the metal by the application of an electric force. The flow of electrons within metallic conductors drives our motors, lights our lamps and heats our kitchen ranges.

The flow of electrons through empty space in a vacuum tube actuates radio tubes, photocells, cathode ray tubes, television camera tubes and electron microscopes. Evidently in these cases, some means must be found always to get the electrons out of the metal and make them travel through the vacuum along predetermined paths.

The most commonly employed method by which electron emission is obtained from a metal is that of heating the metal to a high temperature. That electrons are emitted from an incandescent wire in the presence of an external electrical field was discovered by Thomas Edison in 1881, though not described in our present-day language. Broadcasting transmitting tubes utilize incandescent tungsten filaments as a source of electrons. Wehnelt discovered in 1904 that such metals as calcium, strontium and barium emit electrons more copiously at much lower temperatures than metals do. These "dull emitters," operating at a dull red heat, are used nowadays in all our radio receiving tubes. They consume less power than a bright emitter and consequently are more economical in operation. Dull emitters yield more electrons per dollar.

Electron emission obtained from metals by heat is described in technical language by the term *thermionic emission.* This name was coined at a time when the nature of the emitted particles was still in doubt and it does not describe thermal electron emission clearly. To be sure, there always takes place a definite, though negligible, amount of *ion* emission when electrons are emitted, but the term *thermionic emission* is now used with the

meaning of thermal electron emission, *i.e.*, emission of electrons by heat.

The emission of electrons from a metal surface has, at times, been compared to the evaporation of water vapor or steam from a water surface. At reasonably low temperatures, the water particles cannot escape very easily from the bulk of the liquid since they are held in subjection by the attraction of their fellows. When the water is heated to higher and higher temperatures, the water molecules gain enough kinetic energy to escape from the surface. When the temperature has risen to the boiling point, all the energy put into the water as heat is used to provide escape energy for the molecules at the water surface until all water has evaporated. We picture this process as a mechanical model for thermionic emission. In this case also heat must be provided in order to overcome forces at the surface of the metal which tend to restrain the electrons.

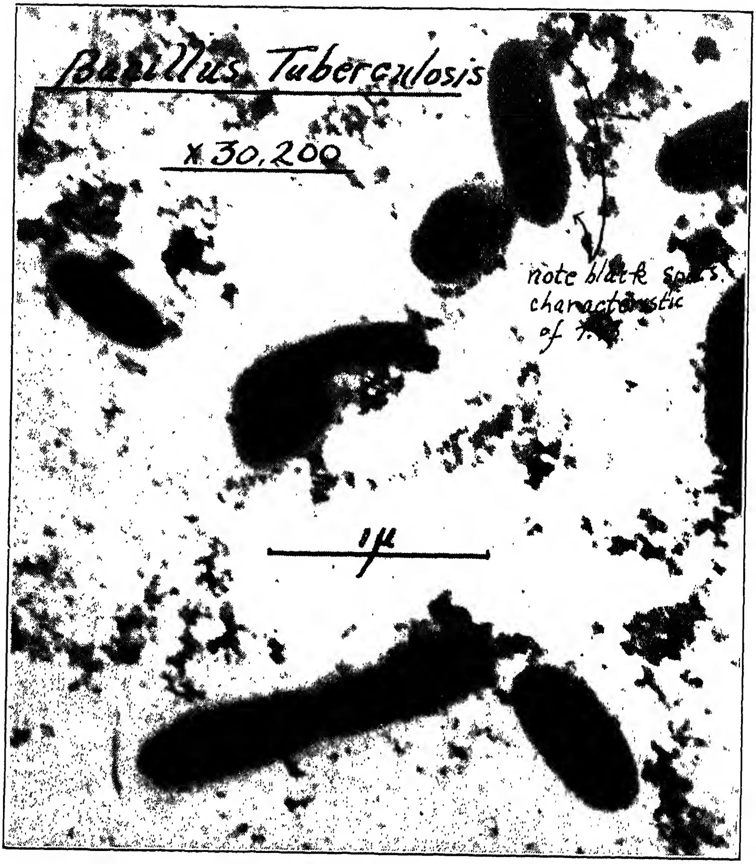
Secondary Electron Emission

We may now utilize the water analogy in explaining another type of electron emission. Small masses of water can be forcibly separated from any large body of water, even at low temperatures, when a foreign body, for instance a stone, is thrown into the water and a splash is produced. The splashing water drops derive their escape energy from the kinetic energy of the stone that has been thrown in. When particles much larger than electrons, such as positive ions, are directed toward a metal surface with sufficient velocity, their impact will "splash" electrons from the metal. Likewise, if electrons of sufficient velocity, obtained from another source, are directed toward a metal surface (or for that matter toward any kind of surface) they will impart their energy to the free electrons of the metal and thus knock them out of the metal. Electrons emitted by the bombardment of a surface with charged or uncharged particles are called *secondary* electrons, and this type of emission is termed *secondary emission.* The bombarding particles, if they are electrons, are called *primary* electrons.

The emission of secondary electrons from metal surfaces, particularly the type used for photocathodes, has found many important applications in recent years. Many electron tubes used in television take advantage of this effect.

Production of X-rays

When electrons impinge upon a metal plate in vacuum they not only create heat and expel secondary electrons but they create an invisible radiation which emanates from the metal and penetrates the glass envelope. This was the phenomenon discovered by Roentgen in 1895, and this invisible radiation became known as x-rays. Experiments in later years have established the fact that x-rays are a special kind of electromagnetic waves (light) of a very high frequency.



Tuberculosis bacillus (×30,200).

Chapter 8

The Dual Theory of Light

Matter and Light

The production of x-rays by electrons impinging on a surface is not only important practically, but very far-reaching theoretically because we have here an example of the energy of matter, electrons, being transformed into light, x-rays. The question presents itself immediately whether the opposite is possible, *i.e.*, to transform light energy into electron energy. That is indeed so. As a matter of fact, this effect was discovered first. During his experiments on the creation of short radio waves from an electric spark, Heinrich Hertz had observed in 1887 that a spark would pass the gap more easily than ordinarily if ultraviolet light fell on the negative pole of the gap. Stimulated by this observation, Wilhelm Hallwachs a year later, found that a negatively charged body loses its charge under the influence of ultraviolet light; this is the so-called Hallwachs effect. In later years, it was shown that a great number of substances, solid, liquid and gaseous, emit electrons under the influence of light. This phenomenon is called the *photoelectric effect.* It is the basis of the modern "magic eye" or photocell. Television pickup cameras are based on the photoelectric effect and so are sound films. Innumerable industrial devices utilize photoelectric cells which are now being manufactured in large quantities. Aside from its practical importance, the discovery of this effect marks a milestone in the history of physical science. It led to the establishment of the dual theory of light. Since this subject is our main concern in this chapter, we must of necessity take pains to study the laws that govern this effect.

The Photoelectric Effect

Let us first become familiar with the physical appearance of a photocell and its operation. Fig. 40 shows the diagram of such a cell in its simplest form. A spherical glass envelope with a long stem, St, is mounted in a base, B, which carries a base pin, P. This pin is connected with a straight wire, A, (or a loop), that is mounted in the center of the sphere on a glass foot. The inside of the glass sphere, with the exception of a window, W, is silvered and electrical connection with the silver film is made

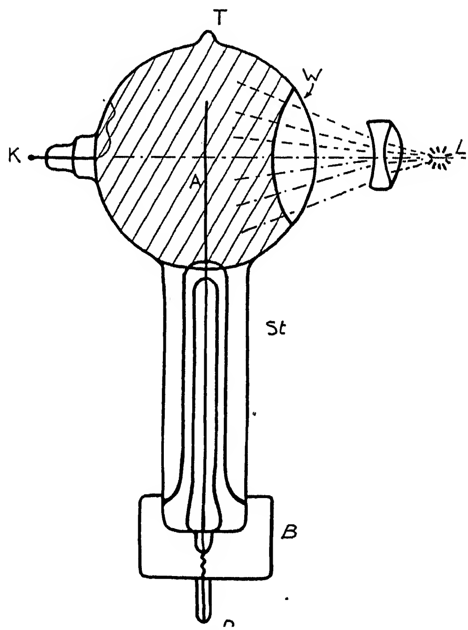


FIG. 40. A photoelectric cell.

through a fine wire which is sealed into the glass and soldered to the cap, K. During the process of manufacture the cell is connected to an exhaust system through a tube attached at T. The silver layer is oxidized by passing a glow discharge in an oxygen atmosphere and then cesium (or some other alkali metal) is deposited on the silver oxide. A carefully controlled baking process permits a reduction of the silver oxide and the formation of cesium oxide with some excess of cesium at the inner surface.

This complex surface, the photo-cathode, has been found to yield a large supply of electrons when it is exposed to a light source, L, sending its rays through the clear window, W. In order to facilitate the flow of electrons through the cell, a positive voltage, E, from a battery is applied to the anode, A; a current, i, indicated by a micro-ammeter, will flow through the circuit as soon as the cell is exposed to light. When the light is shut off, or the window covered, the current falls to zero immediately.

Fig. 41 gives the circuit diagram. A load resistance, R, is included in the circuit. This may take the form of a sensitive relay, which operates any desired device, or the voltage drop across the resistance, R, created by the photo current, i, may be fed to the input stage of an amplifier so that a less sensitive relay may be used in the output.

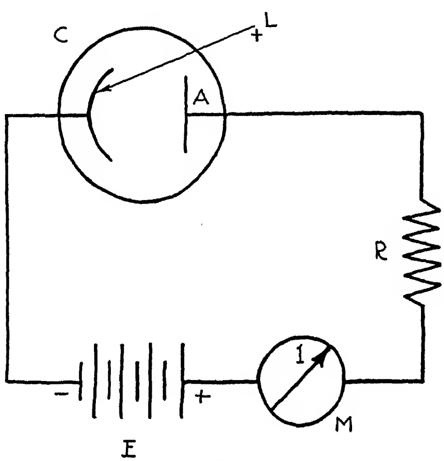


FIG. 41. A photoelectric cell circuit.

The following facts can now be observed:

(1) The photo-electrons are emitted by the cathode, without any measurable delay in time, as soon as the light reaches it.

(2) The number of electrons emitted increases with an increase in the intensity of the light source; *i.e.*, the stronger the illumination of the cathode the greater the number of electrons emitted per unit area of the cathode and the greater will be the deflection of the ammeter.

(3) The frequency of the light source, or, in other words, the color of the light or the wave length of the light, has a decisive effect on the operation of the cell. The *velocity* with which the electrons leave the cathode, or, in other words, the *energy of the electrons* at the moment they leave the cathode, increases with the *frequency* of the light. The intensity of the light has no effect on the value of this initial energy of the electrons. Furthermore, *the frequency of the light source must exceed a certain characteristic minimum value before any electrons are emitted at all.* This threshold frequency, ν_0, or threshold wave length, λ_0, depends on the type of cathode that is used in the cell. It is thus possible to make cells that are particularly sensitive to red light, others that respond to blue or ultraviolet light, and again others that have their maximum sensitivity coincide with that of the human eye at 5,500 A. U. (yellowish green).

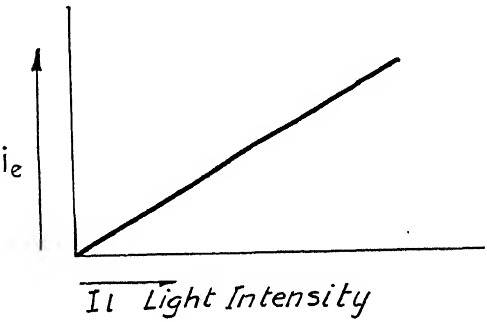


FIG. 42.

Relation between the photoelectric current, i_e, and the intensity of the light causing this current.

The laws governing the photo-electric effect are illustrated by graphs in Figs. 42, 43, and 44. In Fig. 42 the current, i_e, carried by the number of electrons, n, emitted per second per unit of cathode area is plotted as ordinate and the incident light intensity, I_1, as abscissa. We find that i_e is directly proportional to I_1. Fig. 43 shows how the energy, E, of the emitted electrons increases uniformly as the frequency of the light from a constant-energy light source is gradually increased. It is to be noted that a certain definite threshold frequency must be reached before any energy is emitted. Fig. 44 gives the response of a cell to the

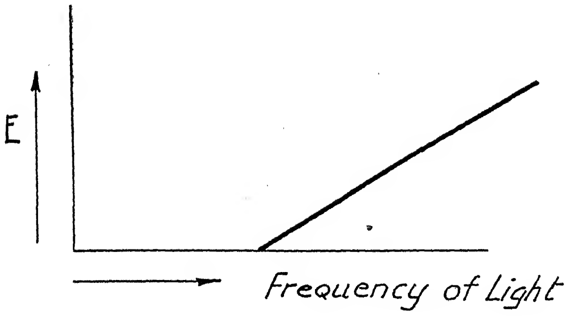


Fig. 43.

Dependence of the energy of the emitted electrons, E, and the frequency of the incident light.

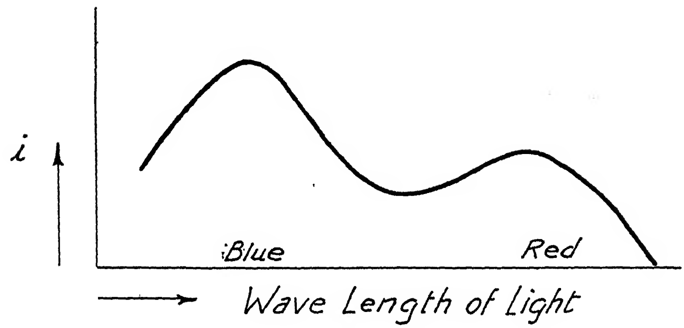


Fig. 44. How the current of a typical photoelectric cell varies as the frequency or wave length of the incident light is changed.

various wave lengths of the spectrum. The photo-current per unit of area which results from a unit of light energy of a given wave length is plotted for the whole spectrum of wave lengths. When the wave length is longer than λ_0 in the infrared region, corresponding to ν_0 in Fig. 43 ($\nu = v/\lambda$), no more electrons are emitted. The example given in Fig. 44 discloses a strong maximum photo-current in the blue region of the spectrum and a lesser second maximum in the red.

The Inadequacy of the Wave Theory

We shall now try to explain these facts on the basis of Maxwell's electromagnetic wave theory of light. Without losing ourselves in too many details, we may apply the familiar concept of any wave theory to the problems at hand. It is, for instance, readily observed in everyday occurrence that a radiating source

gives off energy uniformly in all directions. Thus, a light source can be seen from all directions, a bell on a church tower can be heard from all directions, as long as no obstacle intervenes. The intensity of the light received or of the sound heard will decrease with the distance from the source according to a definite law. It turns out that the received energy falls to one-quarter of its original value when the distance is doubled or to one-ninth when the distance is trebled, and so on: in other words, the received energies are coupled with the distance from the source by the inverse square law. If we keep the distance from the

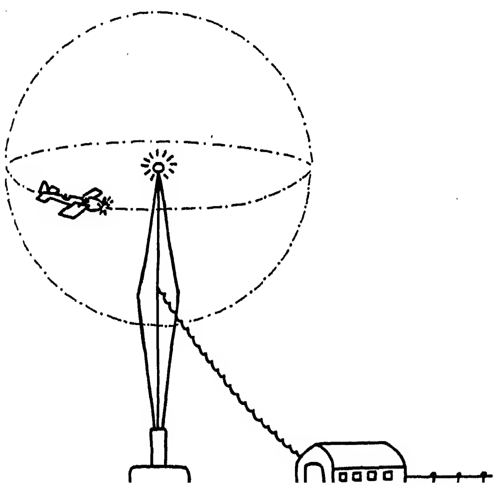


FIG. 45.

How energy received from a point source varies with the distance from the source.

source constant, the received energy at that distance will be the same per unit of time, no matter in which direction from the source the receiver is located. If an airplane flies in a circle around a transmitting tower, the light at its top will appear of a certain constant brightness to the pilot. He could fly on the surface of a sphere with the light source as center and still observe the same brightness; this sphere is indicated in Fig. 45. If the pilot then would fly further away, say, on a sphere twice as large, the beacon would appear to be only one-quarter as bright. If we were to follow the light energy that is sent out by the beacon at a particular instant and trace its propagation through space,

we would have to visualize something like a soap bubble that first snugly fits the light bulb and then expands continuously to enormous proportions. It would of necessity become thinner and thinner as its size increased; the thickness of the film may be taken to measure the energy contained in a unit area of the surface. Since the fixed amount of energy that started to peel itself off from the bulb is spread over a larger area when the size of the sphere has increased, the energy per unit area must of necessity decrease.

When our light energy, or soap bubble—to continue the analogy—reaches a photo-cathode, the surface element of the bubble, say one square centimeter, will impinge on the corresponding surface element of the cathode and so transfer its energy to it. The electrons emitted from this small area by the light energy striking it should then share this energy among themselves and, according to the principle of the conservation of energy, the combined energy of the electrons could not exceed the energy content of the surface element of the light wave. Careful measurements disclose that this condition is not fulfilled. The electrons are actually emitted with a far greater energy than could be conveyed to them by a uniformly expanding spherical wave. It might be suggested that the cathode stores up the energy of a number of successive waves until the required amount is obtained to account for the observed electron energies. This would take considerable time, of the order of hundreds of hours, before the figures would agree. Actually, as has been stated above, no measurable time delay between the arrival of the light wave and the emission of the electrons can be observed.

Our discussion has been based on two assumptions. First, the validity of the principle of the conservation of energy and second the wave nature of light. Since our deductions cannot explain the observed effects one of the assumptions must be wrong. The law of the conservation of energy has been found to be quite consistently true and serves as one of the pillars on which the science of the natural phenomena seems to rest. We must then, once again, call into question the wave nature of light!

The Quantum Theory

The nineteenth century complacency regarding the sufficiency of the wave theory of light received its first shock in 1900 through the work of Professor Max Planck of Berlin. He found it impossible to explain the experimental relations between the absorption and emission of radiant energy by hot bodies and their temperatures by any wave motion model. The classical concept, accepted at this time, was that the continuous flow of energy through space in the form of electromagnetic waves required that both absorption and emission of such energy by hot bodies should also be continuous. But this theory of absorption and emission could not be made to agree with experimental facts, so Planck made the exceedingly bold assumption that, in spite of how the energy is transmitted, absorption and emission at material bodies takes place in very small but definite units, *i.e.*, discontinuously. Each small unit bundle of energy, Planck called a *quantum*; he suggested that there would exist a "kind of atom of mechanical action," which came to be represented by the letter h, now known as Planck's constant.

Now, when the curious results given by the photoelectric experiments were considered, it was apparent that they could be explained if one assumed a corpuscular theory for light. Einstein, in 1905, extended Planck's theory to say that not only are absorption and emission discontinuous but even the energy of a light beam travels through space in quanta. This brings us back to the corpuscular idea of light—Newton's theory!

A light corpuscle, a *quantum* or a *photon*, as it is called, would carry its energy undiminished from the source to the greatest distance and would be able to hand its energy to an electron when it strikes a photo-cathode so that the electron can be emitted instantly. The more intense the light source, the more corpuscles are emitted and the greater the number of electrons set free. This satisfies actual observations.

In order to explain the relation between the electron energy and the frequency of the light in the photoelectric effect (Fig. 43) it becomes necessary to assume that the various corpuscles of light

have different energy contents depending on the frequency of the light. Thus a corpuscle, or quantum, or photon, of blue light contains more energy than one of red light; the electrons set free by blue light move faster than those set free by red light.

Einstein suggested that, as the energy is proportional to the frequency, it might be expressed by the relation:

$$E(\text{the energy}) = (\text{a constant}) \times \nu(\text{the frequency})$$

He further suggested that the constant be put equal to Planck's constant, which is numerically equal to 6.55×10^{-27}. This leads us to the fundamental equation for a quantum:

$$E = h\nu$$

We now find ourselves in a great predicament. All the arguments raised against Newton's corpuscular theory of light must of necessity come up again. We know that a corpuscular theory cannot explain interference, polarization, diffraction. We may recall that Newton himself hesitated to be tied down to the corpuscular theory exclusively and made use of periodic states which imply nothing else but a wave character. This intuitive hesitation of his to accept a one-sided theory is now raised to a basic principle in physics. *Light is neither wave nor corpuscle: it is both.* This is what we mean by a *Dual Theory of Light.*

One should not be frightened at this seemingly contradictory statement. The fact is that we cannot bend nature to suit the niceties of our old-fashioned ideas. Things evidently are not as clear-cut as we would like to have them. We should furthermore bear in mind that the light corpuscle or photon, $h\nu$, is defined by a *frequency*, and thereby is closely tied in with the wave concept. Frequency is a derived term; it means the number of times that a wave recurs per second. To find the value of the frequency, we measure the wave length and then divide it into the velocity of light. We might say that the path of the photon is guided by the wave that is associated with it (Fig. 46). It simply helps our mind to visualize what may be going on. A transverse wave is drawn that weaves itself around the direction

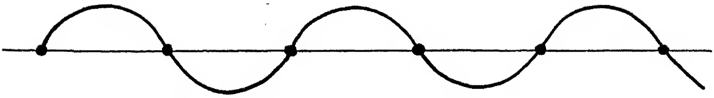


FIG. 46. An electron and its associated wave.

of propagation indicated by the axis. Wherever the wave intersects the axis the heavy dot represents a photon traveling in a straight line.

In disclosing this dual nature of light, not by the choice of their fancy but driven by irrefutable facts, physicists now think of the wave concept when describing phenomena that are evidence of the wave nature of light, and of the photon concept when dealing with the photoelectric effect and radiation problems. In the words of Sir William Bragg: "We teach the wave theory on Mondays, Wednesdays and Fridays and the corpuscular theory on Tuesdays, Thursdays and Saturdays."

When we review the changes that the concepts of the nature of light have undergone during the course of the years and realize the fruitfulness of the concept of duality, we may well turn to the electron and question whether it has an exclusively corpuscular nature. It will be shown in the following chapter that for the electron also a dual theory has been firmly established. The wave nature of the electron is indeed the basis upon which the electron microscope rests.

Chapter 9

The Dual Theory of the Electron

Waves and Matter: de Broglie's Theory

Although the dual theory of light was impressed on the physicist rather early in this century, the far-reaching consequences of the fundamental experimental facts on which it was based were not fully recognized until de Broglie of Paris announced his new views in 1923. We have seen how light must be looked upon as having both a wave nature and a corpuscular nature, in order that we may have a satisfactory model simulating the action found in the various light experiments. In the photoelectric effect we find the energy of light directly inducing motion in electrons. Somewhat later than the time of the discovery of this effect, H. A. Compton (1922) showed that certain experiments could only be satisfactorily explained by saying that the action of radiation, light in the form of x-rays, upon an isolated electron was the same as though there were an ordinary mechanical collision between two corpuscles; the two corpuscles Compton suggested were the photon and the electron.

It was at this point that de Broglie suggested a new concept which is best expressed in his own words:

> "A consideration of these problems led me, in 1923, to the conviction that in the *Theory of Matter*, as in the *Theory of Radiation*, it was essential to consider corpuscles and waves simultaneously, if it were desired to reach a single theory, permitting of the *simultaneous* interpretation of the properties of *Light* and those of *Matter*. It then becomes clear at once that, in order to predict the motion of a corpuscle, it is necessary to construct a *new mechanics*, a theory closely related to that dealing with wave phenomena, and one

in which the motion of a corpuscle is inferred from the motion in space of a wave. In this way there will be, for example, *light corpuscles, photons,* but their motion will be connected with that of Fresnel's *waves,* and will provide an explanation of the phenomena of interference and diffraction. Meanwhile it will no longer be possible to consider the material corpuscles, electrons, and protons, in isolation; it will, on the contrary, have to be assumed in each case that they are accompanied by a wave which is bound up with their own motion. I have even been able to state in advance the wave length of the associated wave belonging to an electron having a given velocity."

We recall that in talking about sound waves, we found a simple relation ($v = \nu\lambda$) between the velocity, v, the wave length, λ, and the frequency, ν. Now, if the energy of the electron is also directly proportional to the frequency, we should expect a simple relation between the velocity of the electron and the wave length to be ascribed to it. This relation is expressed by: The wave length $= \dfrac{\text{Planck's constant, } h}{\text{Mass of electron } (m) \times \text{its velocity, } (v)}$

or $$\lambda = \frac{h}{mv}.$$

Experimental Confirmation of de Broglie's Theory

The theory outlined in the last section was *mere theory.* The important question then was: "Will electrons, or a beam of electrons, exhibit experimentally that there really is a wave motion attached to an electron, or, in other words, will a beam of electrons show the phenomena of interference and diffraction, the *sine qua non* of wave motion?"

After de Broglie's announcement, this question was feverishly worked at; it was answered in the affirmative by two groups of workers, quite independently: by Davisson and Germer (in 1927) of the research laboratories of the Bell Telephone Company, New York, and by G. P. Thomson in work done at Aberdeen (1928). We do not need to go into the details of these experi-

ments, but it will suffice to say that they answered the question decisively. A moving electron has a wave associated with it.

The Wave Length Possessed by an Electron

As noted above, the theoretical value of the wave length that de Broglie worked out for an electron is

$$\lambda = \frac{h}{mv}$$

The value of m, the mass of the electron, a constant for all electrons, was determined first by Sir J. J. Thomson and many times since, so both h and m are well determined constants.

As we shall see later, the velocity of an electron can be easily determined if we know the accelerating voltage (V). Substituting the various experimentally determined values in the above expression, it has been found that

$$\lambda = \frac{12.24}{\sqrt{V}} \times 10^{-8} \text{ cm.}$$

where V is expressed in volts. If the voltage is of the order of hundreds of volts, the wave length is of the order of 10^{-8}cm or less.

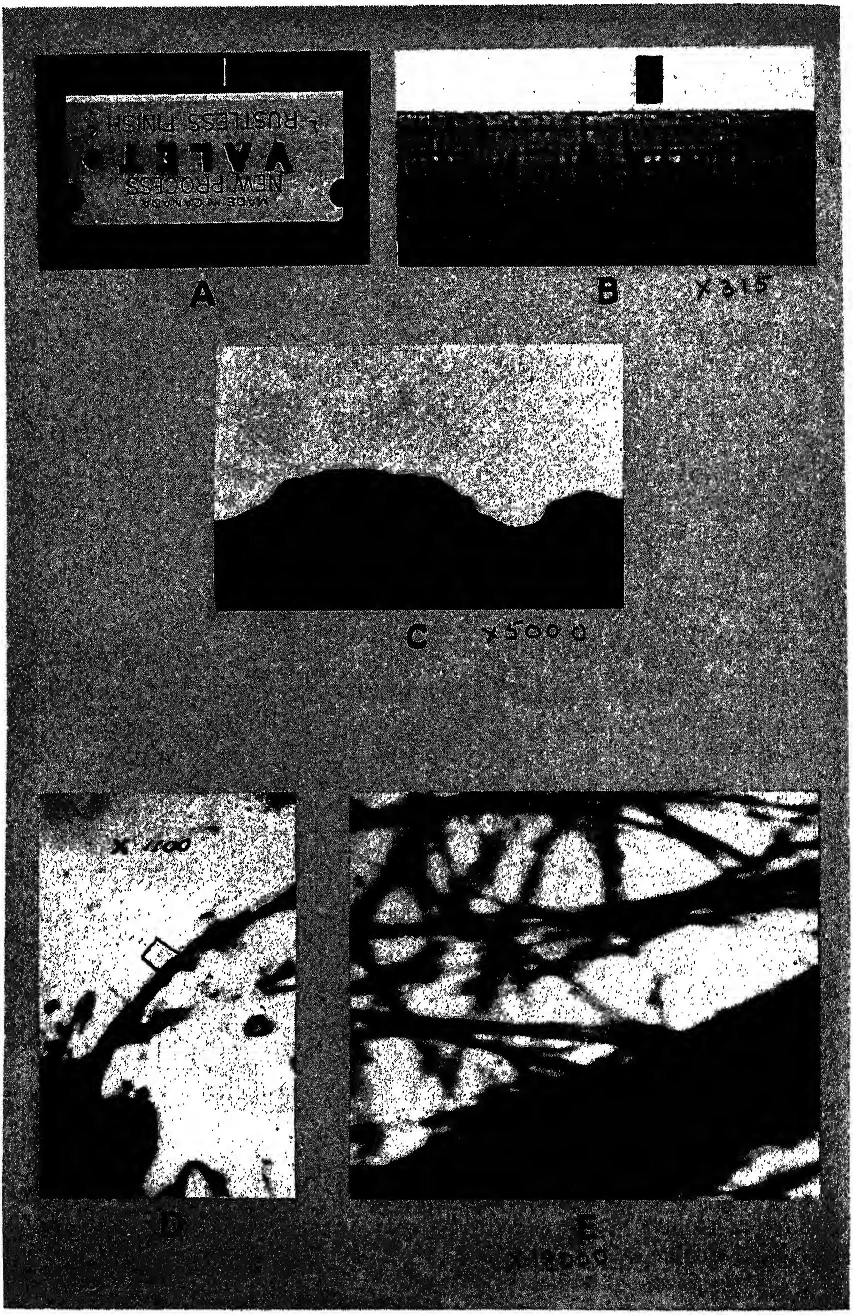
In view of this fact, scientists interested in the microscope at once asked: "Can the electron waves be used in any kind of microscope?" Because, if they could, there would be promise of increasing the resolving power a thousand times or more.

We have seen that we are limited for the light microscope to being able to see particles about $\frac{1}{125,000}$th of an inch but no smaller. (Chap. 2.) We are limited by the wave length of the light. Using electron light, which has very much smaller wave lengths, we should be able to photograph very much smaller particles and also to reveal much greater detail in an object.

This the *electron microscope* has accomplished.

Electron Waves and Electron Rays

In the development of his theory, de Broglie showed that the path of an electron, its so-called trajectory, as it moves through space bears the same relation to its wave property as the *ray of*


MADE IN CANADA
NEW PROCESS
VALET
RUSTLESS FINISH
A
B
X315
C
X5000
D
E

Explanation of plate on page 96

This plate also shows the comparison between pictures taken with the light microscope and the electron microscope.

A. Normal picture of a new razor blade.

B. Light microscope picture of small portion of the same with magnification ×315.

C. Electron microscope picture of a small portion of **B** with total magnification about ×5,000.

The width of the small line above **A** represents the total section included in **B** and the thick block over **B** represents the total width of the part of **B** included in **C**.

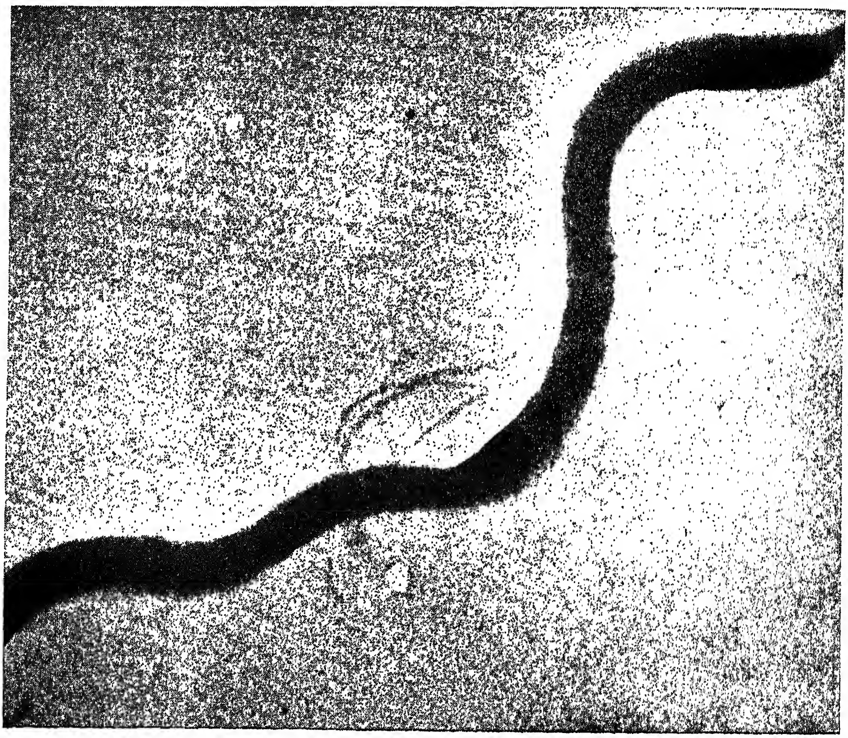
D is an optical picture of asbestos powder (×1,100).

E is the electron picture (×18,000) of the portion outlined by the small inked-in square area in **D**. The heavy bar in **E** is the hazy line through the square in **D**.

These are good illustrations of the fact that the electron microscope reveals structures in a sample entirely unseen in the light microscope.

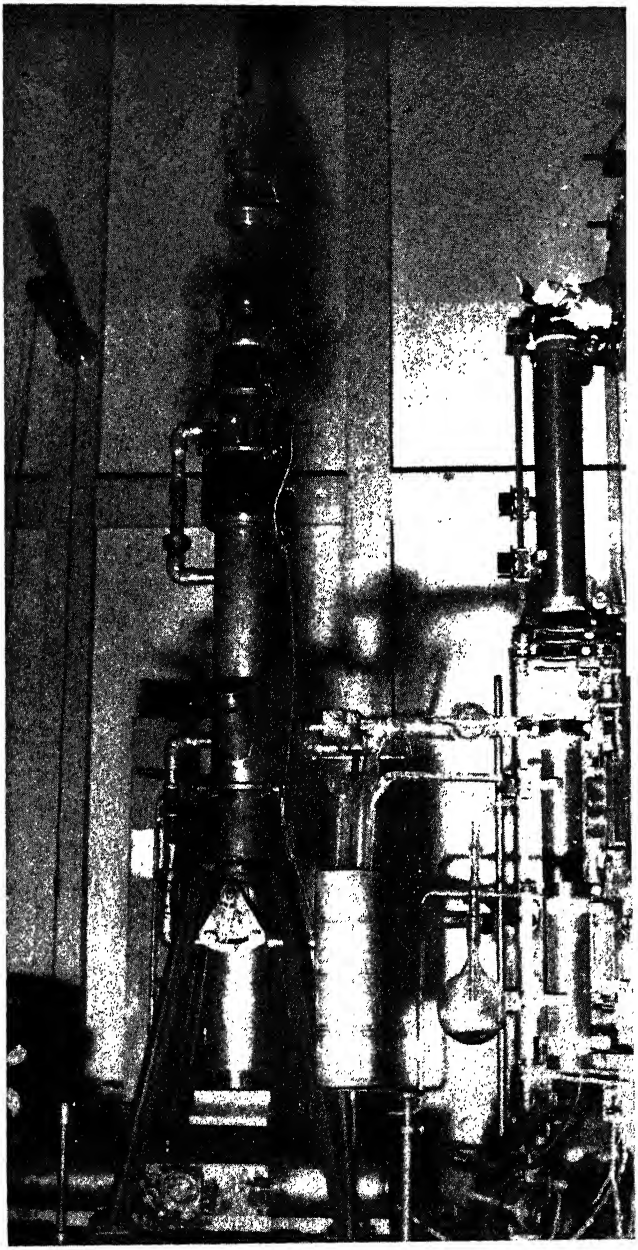
light, which we have used in our early figures, does to the wave property of light. In de Broglie's words, "the possible trajectories of the corpuscle are identical with the possible rays of its wave." Therefore, if we could determine these paths and treat them as "light rays," we might find out something about the possibility of focussing such rays.

There are two ways in which the paths of electrons may be affected: (a) by electrical fields of force and (b) by magnetic fields of force. Of course, these electron beams cannot pass through material lenses, and so if we wish to produce the refraction of electron-waves, we shall have to construct electrostatic or magnetic fields of force, which may perchance serve as *electrostatic lenses* or *magnetic lenses*.



Courtesy RCA Laboratories

Spirochete (*Treponema Pallidum*), the germ of syphilis (×35,000). (Specimen prepared by Dr. K. Polevitzky, Univ. of Pennsylvania.)



The electron microscope in use at University of Toronto. At the right is shown part of the evacuating system and rheostats used to regulate the magnetic fields of the lenses. The overall height is about six feet.

Chapter 10

The Motion of Electrons in Electrical Fields

Work, Force and Power

We come now to a stage in the development of the subject when we shall have to be rather more precise in our language and to introduce some technical terms, which may need some elucidation. Some of these technical terms are simple words in everyday use, which, when used in a strictly scientific sense, need special definition. Such words are *force*, *work* and *power*. In our colloquial use of these terms, we may say a man is "a man of great force" or "a man of great power" and mean the same thing, and probably about the same as to describe the individual as "one who does an enormous amount of work." But when we come to use the term *work*, for example, in the scientific sense, we think of such a thing as carrying a load up a hill; that is exerting ourselves against the force of gravity.

For example, suppose a crate weighing a ton has to be delivered at a factory platform (Fig. 47). It may be hauled up by a block and tackle or dragged up an inclined plane on rollers. It is almost self-evident that the work done or effort expended will be proportional to the pull one has to exert and also to the distance through which we have to pull the crate. It is a matter of common knowledge that we would not have to exert as large a pull if we make use of the inclined plane as we do if we pull the weight straight up. When we make exact measurements of the pull and the distance in the two cases, it comes out that the product of the pull and the distance over which the object pulled moves is always the same for the same height of platform. If we use different lengths of inclines, *i.e.*, inclines making different angles with the ground, we still get the constant value when we multiply pull by distance.

This measure of the work done is just the weight (2000 lbs.) multiplied by the height (20 feet). What we called the *pull* is, in technical terms, called the force exerted; we then have the relation: work done equals force exerted times distance through which the force acts, or in mathematical shorthand simply:

$$W = F \times d$$

In Fig. 47, $F_1 \times d_1 = F_2 \times d_2 = 2000 \times 20 = 40{,}000$ units.

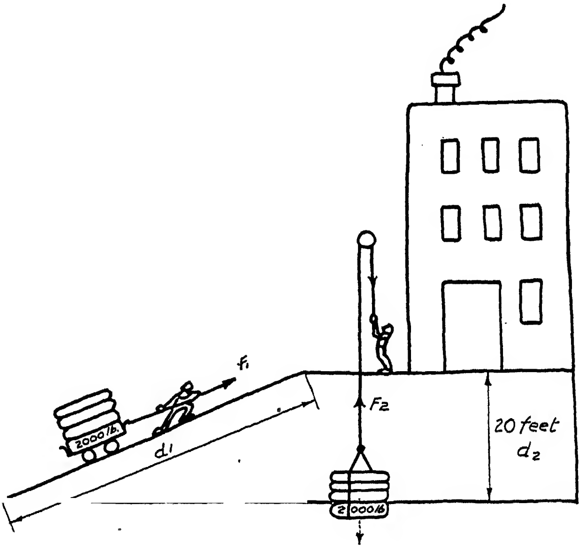


FIG. 47. The work done in pulling the load from the ground to the platform is the same no matter how the work is performed.

We have mentioned also *power*; technically this is just the *rate* at which work is done. The relation is:

$$\text{Power } (P) = \frac{\text{Work done}}{\text{Time taken to do it}} \quad \text{or} \quad P = \frac{W}{}.$$

If in Fig. 47 we pull the rope of the block and tackle by hand, we take a long time to pull the weight up, but if we use a donkey engine of a few horsepower, we can haul up the heavy load in seconds, simply because the power of the engine is greater than the power of one man.

Man's first ideas about electrical charges were that when bodies such as glass or ebonite were put in a state defined by saying they were charged, such bodies exerted a force on each other. As soon as we begin to talk about one charged body exerting a force on another, which is indicated by the bodies being pulled toward or repelled from each other (Fig. 36, p. 70), we think of gravitational forces—the attraction of the earth for bodies free to fall. However, with the electrical forces, we have not only attraction but repulsion. In any case we come to the ideas of forces moving masses, or masses being moved against forces, *i.e.*, work being done; and in the end we have the relations between force, work and power

$$W = F \times d, \text{ and } P = \frac{W}{t}$$

as in the ordinary experiences in the lifting of weights.

In these days we are all familiar with the development of "electrical power," as we say, by harnessing huge waterfalls. When the engineer wishes to calculate the rate of production of energy he uses just the same relation.

Work done in any time equals the weight of the water falling in the time times distance fallen, (d), i.e., $W = F \times d$, since the total weight of the water is just the force of gravity on the water. If the unit of time taken is one second, then the power (P) available at the falls will be:

$$P = \frac{W}{t} = \frac{F \times d}{t}$$

where t is the interval of time in seconds.

Difference of Level and Difference of Potential

When the engineer discovers a new falls or creates a difference of level by a dam, he knows that the power available depends on two things, the quantity of water falling down per second and the height of the falls (or dam). Of course, fundamentally, he measures the height directly by the use of a yardstick or steel tape measure. *But there is another perfectly good way to measure the height, that is, to measure the work done in*

hauling up from the level of the bottom of the falls to the level of the top of the falls a given weight of water, say a pailful.

Here we have just the idea of *difference of level* as referred to the earth; there are two things that a given difference of level between the points A and B will tell us: (1) in what direction water will flow, whether from A to B or from B to A, *i.e.*, we must know which point is at the higher level; and (2) how much work is done in lifting a definite weight from one point to the other. All of these terms have been adopted in the language applied to electricity.

If we have two points, A and B, either in space or on a conductor, say a wire, and there are electrically charged bodies around, which, of course, give rise to electrical forces, then there will be an electrical force acting between A and B and electricity will be moved from A to B (or B to A). This case is more complicated because we have two kinds of electrical charge to talk about, positive and negative, and any electrical forces will, of course, tend to make positive electricity move in one direction and negative electricity move in the opposite direction. So we have to choose one or the other, and a certain charge of positive electricity has been chosen as the standard test quantity; that is, it takes the place of the pail of water used to measure differences of level.

What we call *difference of level* in dealing with water falls is called in electrical language *difference of potential*; this is measured in volts. If we know the difference of potential between two points, A and B, we know which way positive (as well as negative) electricity will move, whether from A to B or B to A, and we know how much work will be done in taking a unit positive charge from one point to the other.

How are we to measure difference of potential? We have no yardstick or measuring tape for it. The only way it is measured is by finding the work done in taking a certain unit quantity of electricity from one point to the other. In fact, *the difference of potential between two points is the work done in taking a standard unit of positive electricity from one point to the other.*

In the case of lifting weights, as in Fig. 47, the work done in lifting a given weight depends *only on the difference of level* between the ground and the platform and not on the path by which the weight is pulled (*i.e.*, straight up or up an incline); so the work done in moving a quantity of electricity from one point to another depends *only on the difference of potential* between the two points and not on the path by which the electricity is moved.

Lines of Force and Equipotential Lines

Fig. 48a represents a specific case in our electrical experiments in which these ideas are used: here we have two parallel metal plates, A and B, which are connected to the ends of a battery so that A bears a positive charge and B a negative charge. If we put our unit positive test charge on a small pith ball, P, it will be pushed away from A and attracted toward B. If we have the experiment carefully arranged, P will move from A to B along lines perpendicular to A and B. These lines then indicate at any point the direction in which the electrical forces are acting. They are called *lines of force.*

Now mark off on these lines points corresponding to equal divisions of the total work done in taking P from A to B along any one of the lines of force reaching from A to B. Four such divisions are indicated by the dots. If we join the corresponding

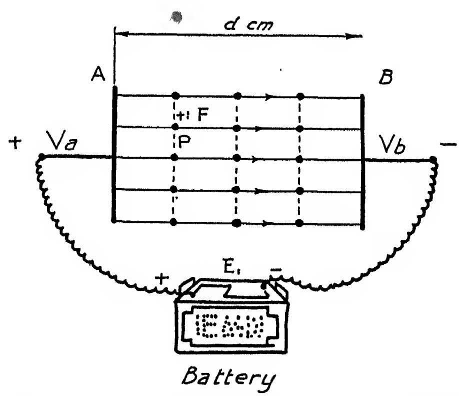


FIG. 48a. Lines of force and lines of equipotential.

dots on the various horizontal lines by broken lines, we get a series of lines which mark off regions of equal work done in going out from A; these vertical lines are known as *equipotential lines*. This is a simple case, and it is apparent that the lines of force are always perpendicular to the equipotential lines wherever they cross one another.

From the symmetry of the lay-out in Fig. 48a, we would easily agree to the statement that the force, F, acting on the small charged pith ball P, will be the same at every point between A and B. From the expression giving the relation between work, force and distance, we can then say that:

$$W \text{ (Work done in taking P from A to B)} = F \times d.$$

But the definition of the difference of potential between A and B is just this work. So if we put the potential of A equal to V_A and that of B, V_B

$$W = V_A - V_B, \text{ and so } V_A - V_B = F \times d$$

From this equation, we get a very useful relation $F = \dfrac{V_A - V_B}{d}$.

Now if everything is symmetrical as in Fig. 48a, $\dfrac{V_A - V_B}{d}$ is the fall of potential per cm., which is consequently called the potential gradient from A to B. We then have the statement that the electric force per unit charge at any point is equal to the potential gradient, sometimes called the gradient of the potential. This quantity—the force per unit charge at any point—is also called the *intensity of the electrical field*, F. It is the difference of potential between A and B divided by the distance between them.

Again, since by definition the difference of potential between two points is the work done in moving a unit positive charge from A to B, we can readily deduce that twice as much work will be done in moving two units of positive charge from A to B and 9 times the original work when a charge of 9 units is moved from A to B. We can express this by the relation:

$$W = q \times (V_A - V_B).$$

Since $W = f \times d$ where f is the electrical force acting on the charge, q, we may write

$$f \times d = q \times (V_A - V_B).$$

Since $\dfrac{V_A - V_B}{d} = F$, it follows that $f = q \times F$, or force equals charge times field intensity. If we put $q=1$, we obtain the definition for the intensity of the field, F, or, as it is sometimes called, the field strength.

If a positive charge q moves from the positive plate A to the negative plate B in Fig. 48a, it moves in the direction of the electric field F, and we say that the field does the amount of work on the charge q. This work appears as the kinetic energy, $\frac{1}{2}mv^2$, which the charged particle of mass, m, possesses when it arrives with the velocity v, at the plate B.

On the other hand, if we were to move a positive charge q from plate B to plate A, the charged particle would have the work W spent upon it. Either we must push the particle along against the force of the field or the particle must possess an initial energy equal to W in order to be able to sacrifice that much on its way from B to A. This can be realized by the following experiment (Fig. 48b).

We arrange a third plate, C, behind the plate B at the same distance, d, which separates A and B and connect another battery, E_2, of the same voltage as E_1, so that plate C is positive with respect to plate B by the same amount that plate A is positive with respect to B. The fields in the space to the left of B and the right of B will then be equal and opposite. If we provide a small hole in B and let the positive charge originate at C, it will acquire the energy W while traveling from C to B and then lose it while travelling from B to A. It will arrive with zero velocity on A.

The same reasoning applies naturally to negative charges. If e stands for the negative charge of an electron, the work done on the electron by moving it from plate B to plate A in Fig. 48a is

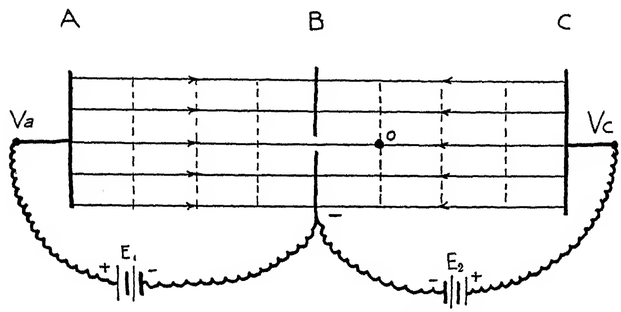


FIG. 48b. Gain and loss of energy of a charge in an electric field.

$$W = e \times (V_A - V_B)$$

and if it were to be moved from A to B, it must have this initial energy in order to be able to overcome the opposing field force. That could be realized by an experiment similar to that illustrated in Fig. 48b. When a field does work on the electron and thus accelerates it, the energy gained by the electron is expressed frequently in terms of *electron-volts*. Thus an electron, which, starting from rest, has been accelerated by moving between two plates having a difference of potential of 50 volts, is said to gain a final kinetic energy of 50 electron-volts of energy. It is also sometimes said to have a final velocity of 50 electron-volts. There are then two kinds of electron-volts, one referring to the energy of an electron, the other to its velocity.

Fig. 49 represents a more general case of lines of force and corresponding lines of equipotential.

We have now to deal with the motions of electrons in vacuum tubes under the influence of electrical forces brought into action by maintaining two metal plates in the vacuum tube at a difference of potential.

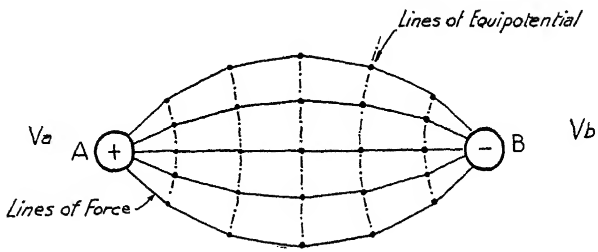


FIG. 49.

The field of force between two conductors at a difference of potential.

The Motion of Electrons in a Uniform Field

The electron stream observed in the earliest experimental tubes such as the Crookes tube and the Braun tube, shown diagrammatically in Fig. 50, was produced by the impact of parts of molecules on the negative electrode. Due to the imperfect vacuum technique used at that time, the most highly exhausted tube still *contained millions of molecules* of various atmospheric gases, and at least a few of these molecules were broken up into positively and negatively charged parts because of a constant radioactive radiation present in space at all times. These positive and negative parts were called *ions*: the negative ion owed its charge to electrons—particles of extremely small mass—while the core of the positive ion was relatively large. These ions would be moved one way or the other in the electrical field inside the tube. The positive ions would be pulled toward the plate C, the cathode, which is negatively charged.

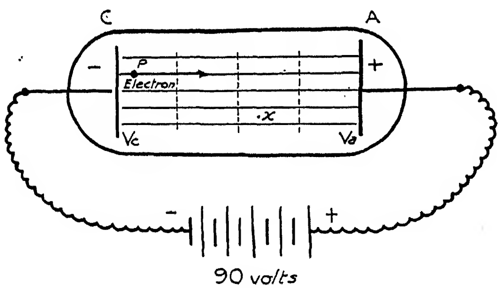


FIG. 50.

The motion of an electron in a uniform field in a vacuum tube.

As we have noted above, these impacts of the positive ion masses will knock electrons out of the plate, C, really in great profusion, and these electrons will be forced away from C toward A. Although many electrons may be knocked out of C per second, we may assume that they are not moving very rapidly just at the surface of C. They then immediately begin to move toward A along one of the lines of force (Fig. 50).

The resulting motion of the electron will be quite similar to the motion of a stone when we drop it to earth. We know that as soon as the stone is released it begins to fall with a velocity

which becomes greater and greater the longer the stone is falling. This is what is called technically "accelerated motion."

Using this analogy we might say that the electron, P, (Fig. 50) "falls" from C to A with a uniformly accelerated motion. As the final velocity with which the stone reaches the earth depends on the height from which it has fallen, so the final velocity of the electron depends on the difference of potential through which it has fallen. The greater the potential difference $V_A - V_C$, the greater is the final velocity.

From what has preceded we see that we always use the term, difference of potential between points, and do not speak of the actual value of anything called potential at any point. This is quite in keeping with what the engineer talks about when he is calculating the power that he can get out of a waterfall. Niagara Falls has a height of 190 feet and the available power is fixed by this *difference of level* but the engineer in making his calculation does not need to know how far the bottom or top of Niagara Falls is above sea level. In fact, sea level is just an adopted level from which measurements of heights of individual stations may be expressed.

Now in dealing with charges on electrical conductors, we are not interested in fixing a value of the potential at one particular point on the conductor, but it is found convenient to define a certain standard *level* of potential. This is the potential of a conductor joined electrically to the earth, or as we say, to ground. With this definition of a standard potential, called zero potential, a positively charged body is said to have a potential above the earth, because positive electricity will flow from such a body to the earth; a negatively charged conductor is one which is said to be below the zero potential, or to be at a negative potential, because positive electricity will flow from the earth to such a body.

Referring to Fig. 50, we may say that the plate A is 90 volts above ground (or +90V) when it is connected with the positive terminal of the battery, the negative terminal of which is connected to a water main, giving good electrical connection with

the ground (the earth). If the positive terminal of the battery is grounded, then the plate, C, is said to be 90 volts below ground (or —90V). The essential point to remember is that we use the term potential only in the sense of difference of potential.

Once more referring to Fig. 50, we may speak of the potential at a point in free space, such as X. This can be found by measuring the amount of work that must be done in order to bring a unit positive charge from plate C to X. If the point X were to lie on the plate C, we would not have to perform any work, since the potential V_C is the same at any point of C. If the point X, on the other hand, were to lie on the plate A, we would have to move a unit positive charge from C through space to A and overcome the repelling electrostatic force all the way until we reach the potential of +90 volts on A. By actually measuring the amount of work performed, it can be shown that it rises linearly all the way from C to A. If the point X lies half way between A and C, the potential at X is then 45 volts above C.

We may now summarize the rules regulating the movement of an electron in an electrostatic field when it starts out from a position of rest. These are as follows:

When an electron is accelerated by an electrostatic field from a position of rest, it moves always along a line of force.

Since the lines of force are always at right angles to the equipotential planes, we may also say that the electron moves at right angles to the equipotential planes when it is accelerated from a position of rest. Consequently if we know either the direction of the lines of force or the lines of equipotential, we can determine the path of the electron.

The Paths of Electrons in Non-Uniform Electric Fields

We shall now consider a case where the equipotential areas in space between cathode and anode are not plane but curved, at least in part. This may be realized by giving the cathode a spherical shape. We know that the surface of a conductor is everywhere at the same potential. If the cathode is grounded, the zero-potential area then must of necessity follow the contour

of the cathode (Fig. 51). The other electrode is again connected to a 90-volt dry battery and the potential values at the various equipotential planes are indicated. Since the anode is a plane, the equipotential areas gradually straighten out the nearer we approach the anode A travelling from the cathode.

We now assume that electrons are emitted with zero initial velocity all over the cathode surface, as indicated by the small dots. From each such electron, a line is drawn toward the anode in such a way that it intersects all equipotential planes at right angles. We know that this is the path the electrons must follow according to rule. The result is that all electrons meet at a small circle, F, at the anode just as if they had been focussed.

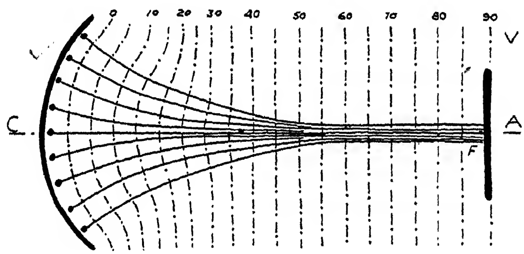


FIG. 51.

Lines of force and lines of equipotential in a non-uniform electric field. As noted in the text, the dotted line marked zero should coincide with the cathode, C.

This sort of thing had been observed by Crookes, Hittorf, Goldstein and others during their early experiments, and they contrived many different shapes for cathodes and anodes in order to learn more about the paths. We must remember, however, that these scientists did not yet know that they were dealing with electrons and they also had no clear conception of potential distribution and its importance for the determination of the paths or trajectories of the emitted particles.

We notice in Fig. 51 that the electrons do not travel in straight lines when the equipotential areas are curved, *i.e.*, when the field is not uniform. In analogy with the reflection of light at a concave mirror one is, however, tempted to speak of an electron focus.

Courtesy American Cyanamid Co.

Zinc oxide smoke (×30,000).

Chapter 11

Electrostatic Electron Mirrors and Lenses

The Reflection of Electrons: Mirrors

One of the simplest or most commonly observed optical effects is that of the reflection of light at a polished surface. Fig. 52 gives a few illustrations of reflection at plane and curved mirrors. Thus Fig. 52a shows a plane mirror, Fig. 52b, a parabolic mirror, with the light source, S, at the focus (as in an automobile headlight), and Fig. 52c a parabolic mirror with a parallel light beam falling upon it. According to the rule that the direction of the light path may be reversed in any optical experiment, the parallel light beam in Fig. 52c comes to a focus at F, and then diffuses as indicated by the arrows.

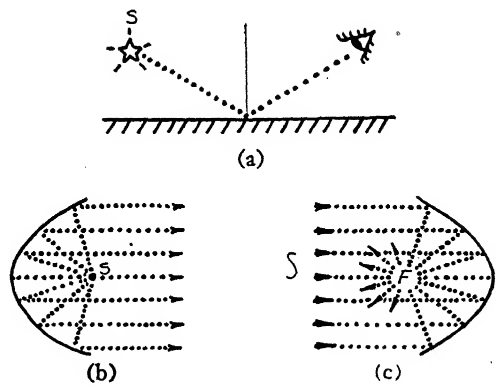


FIG. 52.

Reflection of light at plane and curved surfaces.

These experiments on reflection can be simulated with electrons. We shall treat as an example the case analogous to Fig. 52c. It is assumed in Fig. 53 that electrons, obtained from an emitter some place well to the left, have been accelerated to a velocity of 2,000 volts and move with this velocity within a metal cylinder, A. If this cylinder, A, were closed at both ends,

the potential would be constant and equal to 2,000 volts at any point within. However, if the cylinder, A, has an aperture at its right end, as indicated, and a second electrode, R, of a special shape as shown, is placed near it and charged to -1000 volts, the potential difference between A and R is equal to 3,000 volts. In the space between A and R, where the electrodes are parallel, opposite the flange of the cupped electrode, R, the potential falls nearly uniformly and the equipotential planes are crowded close together and run parallel. Within the aperture of A, however, the equipotential planes bulge toward the left and close to the center of R they naturally follow the curvature of R. For the region near the axis, O, the effect of this potential distribution upon the

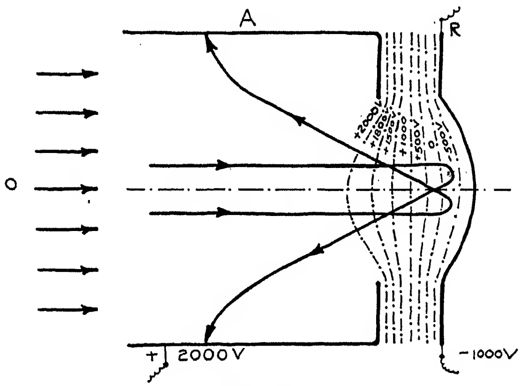


Fig. 53.

electron beam represented by two electron paths is to curl the beam back on itself as indicated. We may thus call the electrode, R, a reflector and speak of the electrode configuration as an electron mirror. This system was described by W. Henneberg and A. Recknagel in 1935.

We note that the electrons do not reach the reflector proper, but are reflected, in the ideal case, at the equipotential plane where the electrons have lost all their kinetic energy and come to rest momentarily, only to be accelerated back toward the left. The electron path is curved, thus differing from the straight path of light beams reflected at a mirror, but the analogy of the electron mirror to the optical mirror is nevertheless striking.

The Refraction of Electrons

It is also evident, from an inspection of Fig. 53, that a bending of the electron path, which was straight within the cylinder A, has been brought about by the potential configuration within and near the aperture. This reminds us of the bending light beam when it passes through media of gradually changing refractive indices. We are thus confronted with the phenomenon of electron refraction.

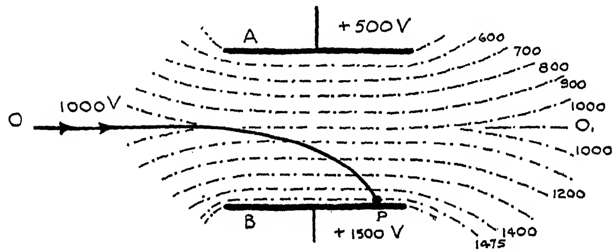


FIG. 54. How the path of an electron bends when travelling through a potential field.

In Fig. 54 the potential distribution between two plates, A and B, is shown for the case where plate A is kept at a potential of +500 volts and B at a potential of +1500 volts. The areas of equipotential will be distributed symmetrically as marked.

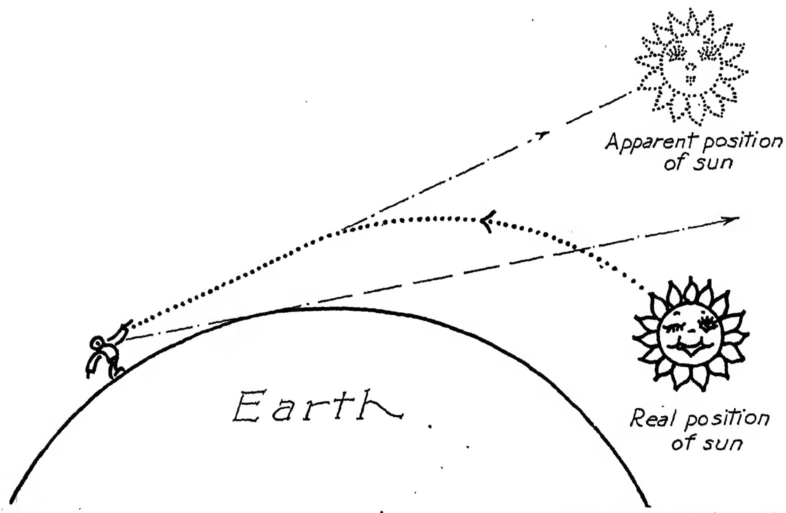


FIG. 2 repeated to show clearly the analogy between the refraction of light and the bending of the path of electrons.

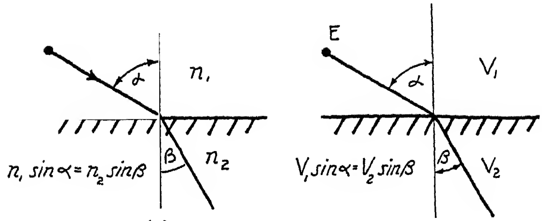
If an electron beam with a velocity equivalent to 1,000 volts enters along the axis O-O_1 from the left, it will follow a curved path as shown and strike the plate B at P, if the potential of B is high enough, or the plate sufficiently long. The layers of increasing potential thus act upon the electron beam in much the same way as the layers of differing refractive index act upon the light beam. Fig. 2 is repeated here in order to bring out the analogy that exists between the refraction of light in the layers of the atmosphere of different densities and this electron refraction.

Electron Optics

Thus far, we have spoken in terms of analogies and emphasized the similar behavior of light beams and electron beams under certain conditions. We must now ask ourselves whether or not a more far-reaching analogy exists that can be expressed in mathematical language. In other words, do the laws of optics apply to electrons?

That this is indeed the case was first recognized by H. Bethe in 1927, who applied Snell's law of refraction to electrons that are refracted at the boundary of two potential layers. It is only necessary to replace the refractive indices, n, in Snell's law by the square root of the potentials, V. Since the electron velocity, v, is proportional to $\sqrt{V}$, we can replace n by v and obtain the two expressions given in Fig. 55, where light refraction and electron refraction at a plane boundary are illustrated. Fig. 55b is an idealized diagram since a sudden jump of the potential value in free space from V_1 to V_2 at the boundary is assumed. This condition cannot be realized in practice as we have seen in previous illustrations. The electron path always bends gradually, and does not show a kink as in Fig. 55b, because the potentials vary continuously and not discontinuously. This does not affect the validity of our deductions.

We stand now at the gateway that leads into an entirely new field of exploration. We need only remember that glass lenses and prisms are based on the phenomenon of refraction, in order to realize what possibilities lie ahead. It will be possible to form



(a) (b)

FIG. 55.

The path of a ray of light going from a medium of refractive index n_1 into a medium of index n_2; Snell's law is that $n_1 \sin \alpha = n_2 \sin \beta$.

The analogous illustration for the path of an electron going from a region of one potential to a region of another potential. Bethe's law is that $V_1 \sin \alpha = V_2 \sin \beta$, where V_1 and V_2 are the velocities of the electron in the two regions respectively.

electron lenses and electron prisms, not of glass, of course, but of the intangible and therefore easily penetrated quantity, potential. We must shape equipotential areas so as to approximate the form of familiar lenses in optics. How this can be done is evident from Fig. 53. There the equipotential areas protrude into the aperture of A with an approximately spherical shape. The geometrical form of the electrodes to which given potentials are applied determines the potential field in free space. Apertures are likely to be a very useful means for obtaining such bulging potential areas. Let us then look a little more closely at some of the electrode configurations that can be built up from apertures and study their effect upon electron beams.

Electrostatic Lenses for Electron Beams

If we arrange three metal disks A, B, C, in a vacuum tube and assign to them the potentials as shown in Fig. 56a, we may plot the potential values that exist along the axis O-O_1 of the system. Disks A and B are at zero potential and disk C is at 100 volts. The *space* between A and B is thus at zero potential, if the stray fields at the edges are neglected, and the potential plotted in Fig.

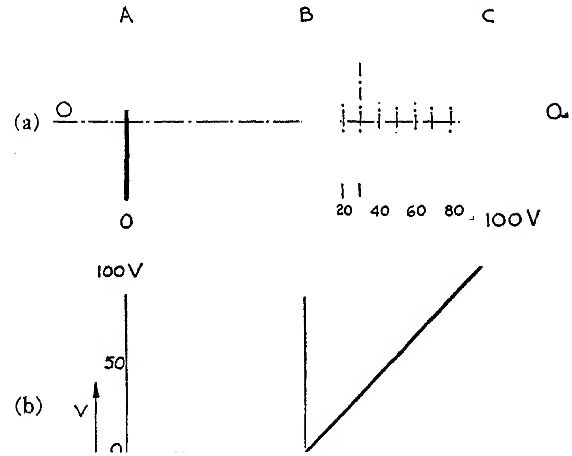


FIG. 56. Distribution of the electrical field between conducting plates at different potentials.

56b along the axis, Z, remains zero. In the space between B and C a potential gradient exists. The field is constant and has a definite positive value. This is indicated by the equidistant equipotential planes, represented by the thin vertical lines between B and C in Fig. 56a, and accordingly a linear rise of the potential plot in Fig. 56b along the axis between B and C.

In Fig. 57a the same electrode arrangement is shown, but the disk, B, is provided with an aperture. The field to the right of B now bulges through the hole in B into the space to the left of B where previously no field existed. The equipotential planes become curved in the vicinity of the aperture and straighten out only in front of plate C. The potential values are written onto the potential areas in small figures. Fig. 57b gives the corresponding rise of potential along the axis. This plot is no more a straight line throughout, but contains only a short straight section in front of A. The slope of the potential plot is small, *i.e.*, the field is weak, to the left of B and it increases rapidly to the right of B. It is worthy of note that the potential at the center of the aperture is greater than the potential of the disk itself.

Let us now assume (Fig. 58) that disk A is an electron emitter at the center. This is readily realized in practice by inserting

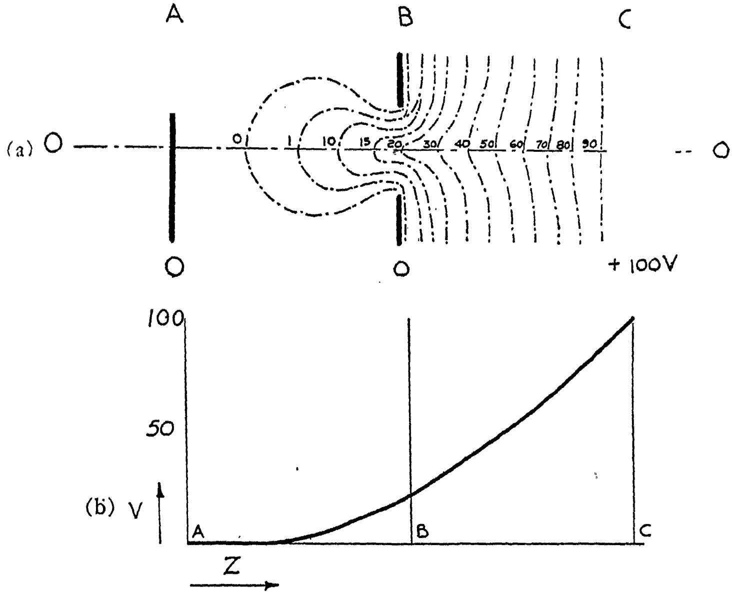


FIG. 57. How the potential distribution is affected by an aperture in an electrode.

the oxide-coated face of a cathode sleeve at the center; this is heated by a flat spiral heater wire back of it. The cathode sleeve, generally made of nickel, is connected with the main disk A by a thin wire, *w*, so as to maintain zero potential on both of them. The disk A thus extends the zero potential face of the cathode without draining any noticeable amount of heat from it. This

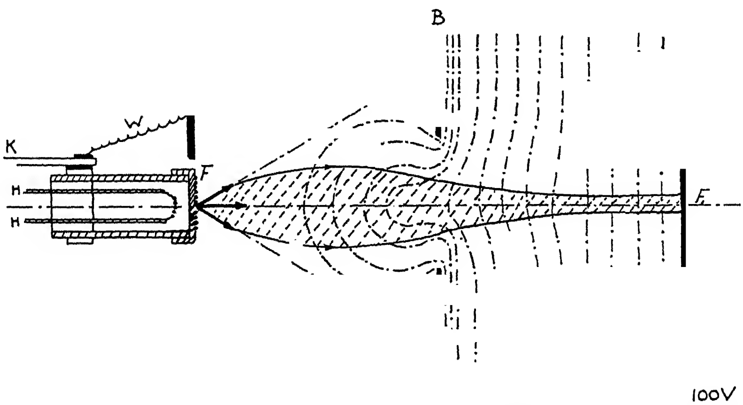


FIG. 58. The influence of an electrode with an aperture near the cathode.

arrangement is frequently used in cathode ray tube practice and is known as a guard ring. When a current is sent through the heater wire, the oxide-coated face, F, will reach a dull red temperature and electrons will be emitted from it. The paths of electrons near the axis are shown brought to a focus, F_1.

Electrons Emitted with an Initial Velocity

We have previously assumed that electrons are released at the cathode surface with zero velocity. This is true for some electrons; others possess an initial velocity even in the absence of an electric field in front of the cathode. Furthermore, while most of the emitted electrons leave the cathode face at right angles in the direction of the Z-axis, others are emitted at an angle with respect to the normal. This is indicated by the three small arrows in Fig. 58 emanating from the center of F. We may illustrate this space distribution of the electrons emitted from a point, P, at the cathode, C, in Fig. 59. Here a circle is drawn in front of the cathode face and a number of arrows are inscribed which all originate at P. The lengths of these arrows indicate the relative number of electrons that are emitted in the direction given by the arrows. It becomes clear from this diagram that many electrons are emitted in a direction normal to the cathode face, but a definite fraction of the total emission occurs sideways.

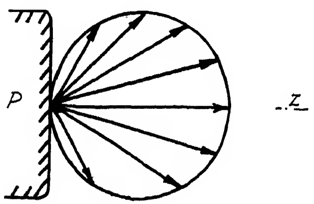


FIG. 59.

The number of electrons emitted from a point on the cathode with velocities in different directions.

It may then be stated that electrons are emitted from a cathode with different initial velocities and in different directions. Another noteworthy feature of the movement of electrons in an electric field must now be explained before the electron paths in Fig. 58 can be fully understood. It was stated earlier that elec-

trons which start from a position of rest travel in an electric field in such a manner as to follow the field lines or, what amounts to the same thing, so as to penetrate the equipotential planes at right angles. This was the case in Fig. 51, where we assumed an emission with zero initial velocity.

We have learned in the meantime that the initial velocities may have positive values, up to about 0.2 volt to be specific, and that these initial velocities may be directed sideways with respect to the normal to the cathode face. How will this effect the electron paths? A little earlier, the rectilinear movement of an electron in a uniform electric field from a position at rest was compared

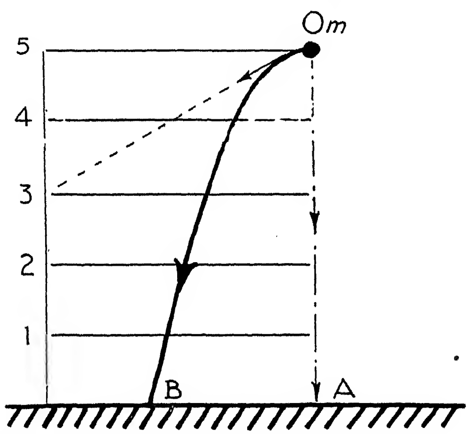


FIG. 60a. The path of a falling stone which has been thrown out with a velocity in the direction of the arrow.

with a stone that falls to earth in a straight line in the gravitational field when released from a position at rest in our hand. If we now give the stone an initial velocity that is directed at an angle to the previous path, the trajectory of the stone will be a parabola, as shown in Fig. 60a. The horizontal lines 1, 2, 3, 4, 5 represent equipotential planes, *i.e.*, points where the work that must be done to raise the stone from the ground level to the height indicated is constant. We note from Fig. 60a that the mass m moves at right angles to the equipotential planes from O to A, when it starts either from a position at rest or when its initial velocity is directed at right angles to the equipotential planes. On

the other hand, when the initial velocity is directed sideways, the path O-B no more follows this rule. The mass, *m*, is then refracted at the respective equipotential planes.

The same reasoning applies to an electron which moves in a uniform or non-uniform electric field with an initial velocity which makes an angle with the field lines. This is illustrated in Fig. 60b. The electron trajectories in Fig. 58 are now readily understood. The electrons which are emitted in the direction along the axis travel in a straight line to F_1 on plate C, just as the axial ray in an optical system reaches the focus in a straight line. The focussing action of the potential field is exerted on the "stray electrons" which are emitted sideways at angles not too large.

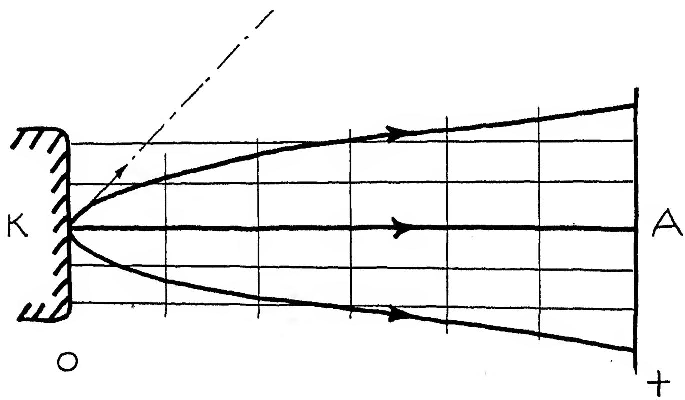


FIG. 60b. The paths of electrons emitted with a velocity making an angle with the direction of the lines of force of the electrical field. This explains directions indicated in Fig. 58.

The focussing action of an aperture system upon an electron stream was first recognized and expressed in mathematical form by Davisson and Calbick of the Bell Telephone Research Laboratories (New York) in 1931. These workers published the formula for the focal length of *an electrostatic electron lens* of the aperture type expressed in terms of the voltages applied to the electrodes. The ground work in this new field of electron optics had been done a little earlier in Germany by H. Busch in 1926. He was able to establish the mathematical theory of the magnetic electron lens of which more will be said in Chapter 12.

Chapter 12

Magnetic Lenses

Magnets and Magnetic Fields

Everyone is familiar with the ordinary magnetic compass—a small piece of steel pivoted so as to move in a horizontal plane. Its usefulness depends on the fact that one particular end always points toward the north, when the compass is kept at rest and level. The end which points to the north is called the north pole, that pointing to the south, the south pole (Fig. 61).

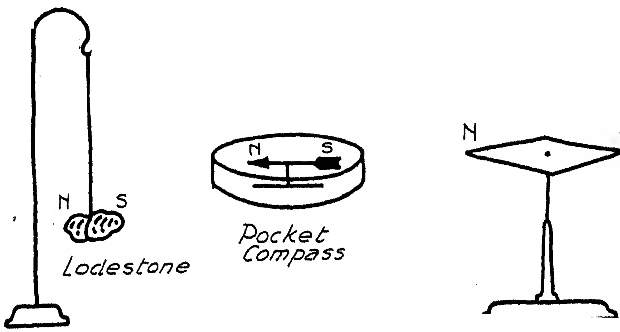


Fig. 61. Lodestone and compass needles.

It is really not known exactly when compasses were first used, but the earliest forms use a magnetic substance which occurs naturally on the earth's surface. This substance called lodestone (leading-stone) is made up of oxides of iron. If an irregular piece of this mineral is suspended by a thread so as to be balanced horizontally, the longest dimension will always set so as to point in a north-south direction—one definite end always pointing to the north.

The question may be asked, "How did the modern small steel magnet develop from the lodestone?" This leads us to speak of

a very important property of lodestone, *i.e.*, that it is able to attract to itself pieces of iron or substances containing iron. Not only will the lodestone attract iron, but if it comes in contact with a piece of iron, it will impart its own power to the iron and so convert it into a magnet. Every boy with a jack-knife knows he can magnetize it by rubbing it on a magnet. All our magnets are made of steel, because steel will retain this curious magnetic property longer than soft iron.

When people began to experiment with these magnetic needles, they discovered a very curious property which has to do with the interaction of one magnet on another. If we have two

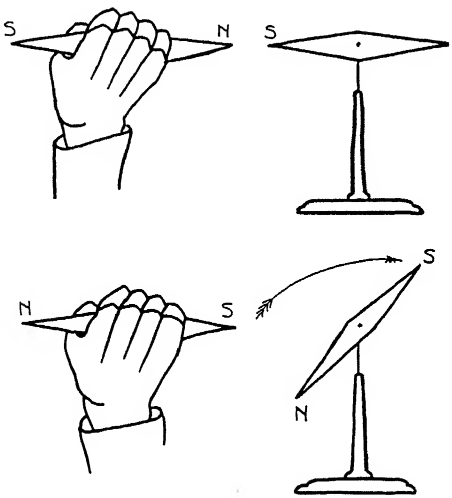


FIG. 62.

The laws of magnets; unlike poles attract each other, like poles repel each other.

magnetic needles and pick one up and bring one end up to the end of a second magnet, we find that two poles which are different (N and S) always attract each other, but that two poles of the same kind (N and N, or S and S) always will repel each other (Fig. 62). This last fact recalls to us something similar regarding electrical charges—like charges repel, unlike charges attract each other—but this does not prove any real connection between magnets and electrical charges.

One very remarkable difference is apparent at once. We may impart to a body a positive or a negative charge and detect its

presence by simple tests. But we cannot detect both a negative and a positive charge on an isolated conducting surface, since the two charges neutralize each other. For magnets, the opposite condition exists. Each magnet always has a north pole and a south pole and we can never get any separate portion of a magnetic substance with only one kind of pole, north or south. If we have a piece of steel which is magnetized and break it into any number of parts, all the parts are still perfect magnets with both north and south poles (Fig. 63).

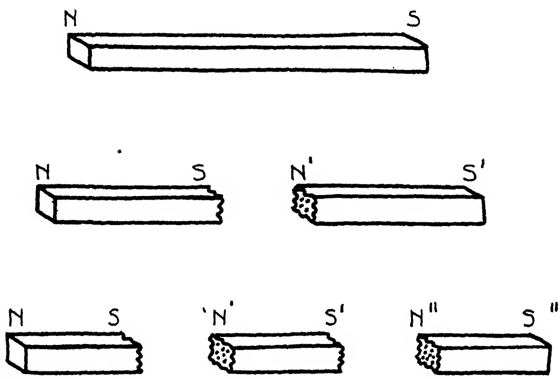


FIG. 63. Breaking up a magnet into many parts always makes a complete magnet out of each part.

Since one magnet will exert a directive force on another, and since any magnet always sets itself in a north-south direction when free to do so, we decide that the earth itself must have a directive action on a magnet. From this we conclude that the earth itself is a huge magnet also having two poles, *i.e.*, points toward which the north and the south poles of magnets are directed. The magnetic pole in the northern hemisphere is in northern Canada near Hudson Bay; this point is known as the *north magnetic pole.* Neither of the earth's magnetic poles are in the same place as the geographical north and south poles.

Just as in the case of the attractions and repulsions between electrical charges, so we have here the remarkable fact that one body affects another apparently "at a distance." In this case also Faraday ascribed this interaction to something going on in the

space round about the magnets. We speak, then, of the region round about a magnet as a *magnetic field of force*, and also of magnetic *lines* of force which show us graphically the direction of the magnetic force at any point round about any magnet. These magnetic lines of force can be easily found in a plane by moving a small compass around on a piece of cardboard placed over the magnet and marking at every point the direction in which the compass needle sets itself.

Fig. 64a, b shows the distribution of lines of force about a bar magnet and between the poles of an ordinary horseshoe magnet;

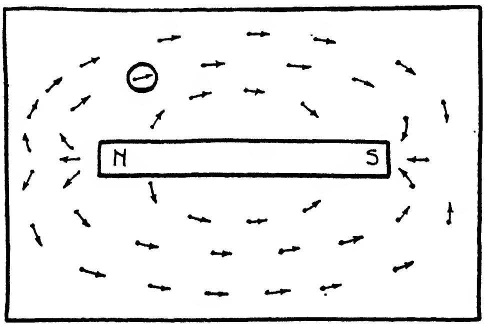


FIG. 64a.

The magnetic lines of force about a bar magnet.

the remarkable and, as we shall find later, useful property of the field of the horseshoe magnet is that the most intense force is in the region between the poles and that in the central part of this region the field is uniform, *i.e.*, the force on a unit pole is the same at every point. (Fig. 64b).

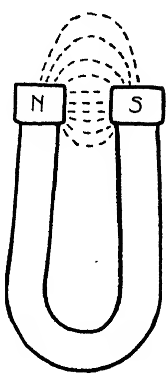


FIG. 64b.

The magnetic lines of force for a horseshoe magnet.

It has been agreed to define the direction of the line of force as the direction in which a north pole tends to point. This is the direction indicated by the arrows in Fig. 64a.

The Magnetic Field about a Conductor bearing a Current of Electricity: The Electromagnet

What has a magnet got to do with an electron? The answer to this question goes back to an accidental discovery in 1820 by a Danish scientist, Oersted, that an electric current flowing through a wire exerted a magnetic effect in the region round about the wire. In other words, the current through the wire set up a magnetic field. This is illustrated in Fig. 65. If we follow the

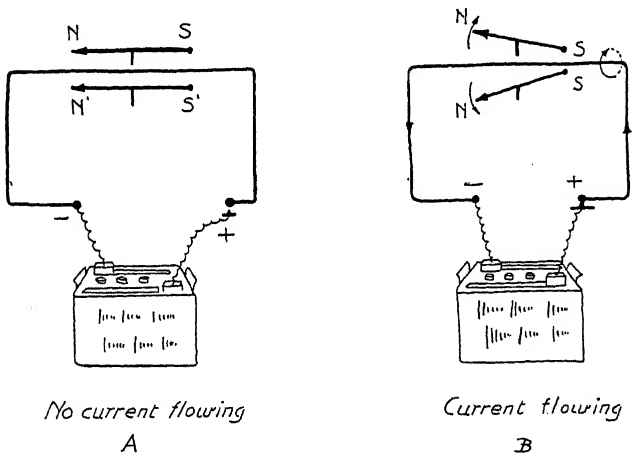


FIG. 65. The direction of the magnetic lines of force about a conductor through which a current is flowing.

movement of the poles of the magnet in its various positions, we are led to the conclusion that the magnetic lines of force about the wire are circles in planes at right angles to the wire and with their centers in the wire. The relation between the direction of the current and the direction of the magnetic lines of force is given by a very simple rule (Fig. 66). If we are using a screwdriver to drive in a screw, and let the direction of progress of the screw represent the direction of the current, the direction of

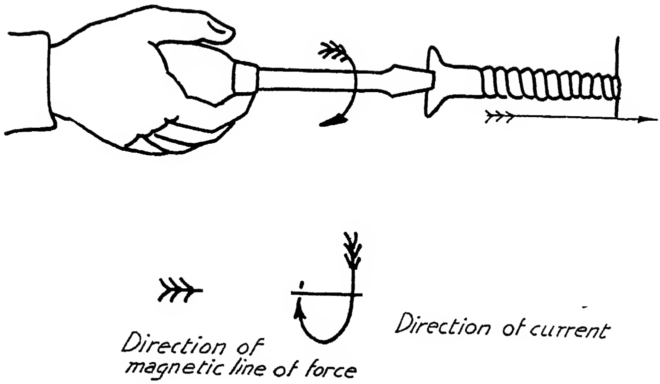


FIG. 66. A simple rule for remembering the relation between the direction in which a current is flowing in a conductor and the direction of the magnetic field set up.

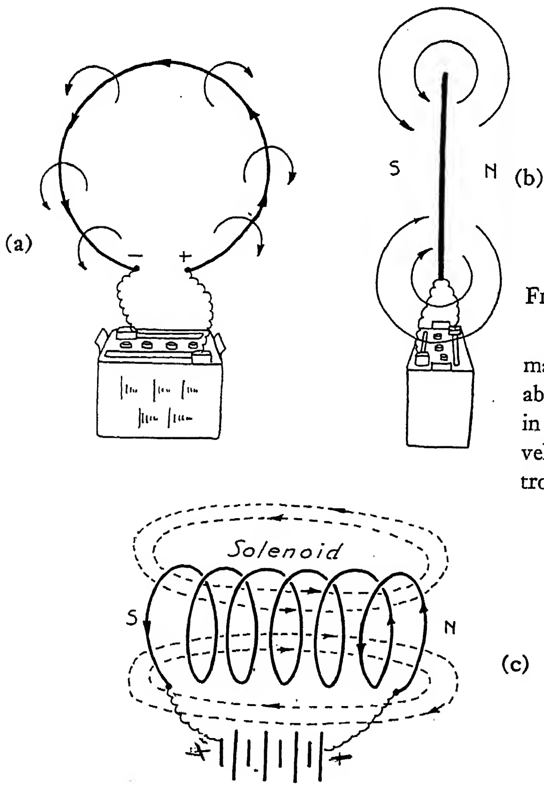


FIG. 67.

The direction of the magnetic lines of force about a conductor bent in a circle and the development of the electromagnet.

rotation of the thumb represents the direction of the magnetic line of force.

Now if we bend the wire of Fig. 66 in the form of a circle, as in Fig. 67a, the lines of force, if we could see them, would appear to come out of one face of the circle, bend around the wire and re-enter the other face of the circle of wire. This is shown in the plane of the circle of wire in Fig. 67b. In fact this figure suggests that the circle of wire acts as though it were a very thin flat magnet, such as we might get if we sliced out a very thin section of the bar magnet shown in Fig. 64a.

If, instead of a circle of one turn, we have a long wire coiled up in what is called a *solenoid* (Fig. 67c), we have an exact parallel of a long bar magnet; such a solenoid is known as an electromagnet. One end of the solenoid is a north pole and the other end a south pole, the polarity of either end depending on the direction of the current through the coil.

In this case we have produced the equivalent of a bar magnet without the use of any lodestone, iron or any other magnetic substance. One very important property of such an electromagnet is that the strength of the magnetic field within the coil of wire is *uniform* (see Fig. 68). However, we can get a much

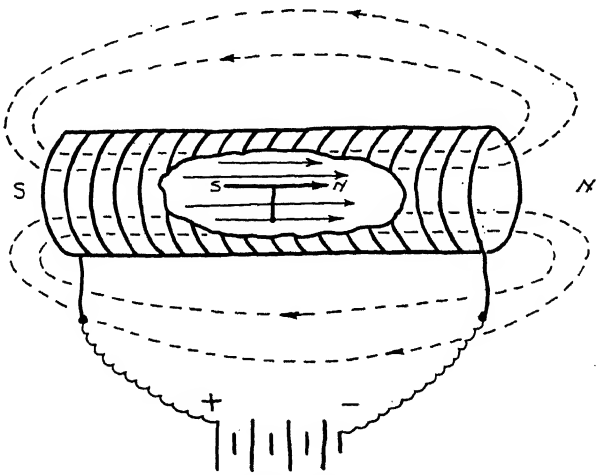


Fig. 68. The magnetic field at the center of a long solenoid is uniform.

stronger magnetic field about the electromagnet if we insert a core of iron or steel in the solenoid.

The great usefulness of the electromagnet is due partly to the fact that by increasing the value of the current through the wire, we can increase the strength of the magnetic field over a great range of values. Again, by employing an alternating current, it is possible to cause a corresponding variation in the direction of the magnetic force.

The Action of a Magnet on a Conductor Bearing a Current

Oersted's discovery showed that a current in a conductor can move a magnetic needle: that is, a fixed conductor bearing a current affects a movable magnet. The converse of this is that a fixed magnet ought to produce motion of a movable conductor which is bearing a current. This can be shown to take place in a simple experiment (Fig. 69). A battery sends a current through

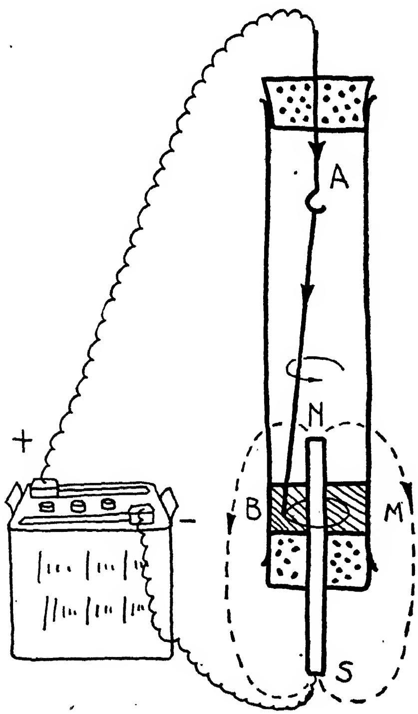


FIG. 69.

Showing that a stationary magnet can produce motion in a conductor bearing a current; the converse of the action shown in Fig. 65.

a wire, AB, which is suspended from a metal hook, A, and dips into a layer of mercury, M; a bar magnet is supported through a cork at the bottom of the tube. When the current is sent through AB, the lower end rotates about the pole of the magnet as long as the current flows. Here the motion of the conductor is perpendicular to both the direction of the current and the direction of the magnetic lines of force; in the case of Oersted's experiment, the motion of the magnetic pole is in a direction perpendicular to both the direction of the current and the direction of the magnetic lines of force from the magnetic needle.

Now if electrons are negatively charged particles, a stream of electrons moving along a straight line should be equivalent to a flow of negative electricity in that direction. Consequently, if the stream of electrons is sent through a magnetic field, the magnetic force should move the electrons in the stream in a direction at right angles to both the direction of the electron stream and the direction of the magnetic field (Fig. 70).

It was just this observation which led Sir J. J. Thomson to conclude that the electron stream acted as though it were a stream of negative electricity and to announce that electrons were small particles of matter charged negatively.

Experiments show that the electron is affected by a magnetic field only when it is moving. If the electron is at rest, there is no action whatever.

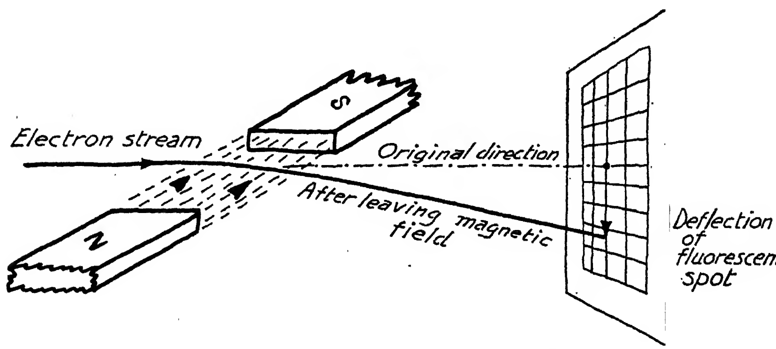


FIG. 70. The deflection produced in the path of an electron passing through a magnetic field.

Comparison of the Action of Electric and Magnetic Fields on Electrons

We find then that the flight of electrons is affected by the presence of a magnetic field as well as by the presence of an electrical field. It is important to contrast the effects of these fields.

An Electrical Field

1. Acts on an electron which is initially at rest and causes it to move along the lines of force in a direction opposite to the motion of a positive charge.
2. Acts on moving electrons in such a way as to bend the beam into regions of higher potential (Fig. 55). This refraction of the electron beam at the equipotential boundary layers is accompanied by a variation in the speed of the electrons.

A Magnetic Field

1. Has no effect whatever on an electron which is at rest.
2. (*a*) Has no effect on a moving electron as long as the direction of motion is along the lines of force of the magnetic field.
   (*b*) Moves the electron in a circular path, if the direction of motion of the electron is at right angles to the lines of force of the magnetic field. This circular path is still perpendicular to the magnetic lines and the speed of the electron is not altered.

Fig. 71a illustrates what will happen to a moving electron entering a uniform magnetic field in a direction at right angles to the lines of force. Immediately the electrons enter the magnetic field, a force is exerted on them which tends to bend the

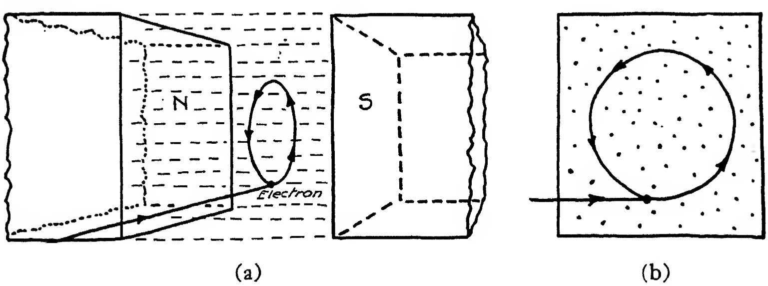


Fig. 71. The path of an electron entering a magnetic field in a direction at right angles to the lines of force is changed to a circle. Fig. 71b gives an end-on view of the circular path; the dots represent the sections of the lines of force.

electron beam into a circle of definite radius; this circular motion persists as long as the electron remains in the field. The value of this radius, R, varies directly as the velocity of the electrons, and inversely as the intensity of the magnetic field. The technical expression is $R=\frac{mv}{eH}$, where m is the mass of an electron, v its velocity, e its charge, and H the value of the magnetic field; these values must be expressed in terms of definite units in order to get the value of R in centimeters.

If the magnetic field occupies a small space, the electron may escape from the region of the field before it completes the circular path; this is illustrated in Fig. 72. The electron beam enters the field at A and escapes at B; while moving in the field, the electrons follow the path AB, which is an arc of a circle with the radius given by the above relation. That is, AB is a part of the dotted circle, which would really be the path of the electron if the uniform magnetic field extended farther in space.

At the point B, where the electrons leave the magnetic field, the action of the magnetic force ceases and the electrons continue their flight in a straight line along the tangent of the circle at B. The circle which is indicated by the dotted line in Fig. 72 appears as an ellipse since it is drawn in perspective. Figs. 71b and

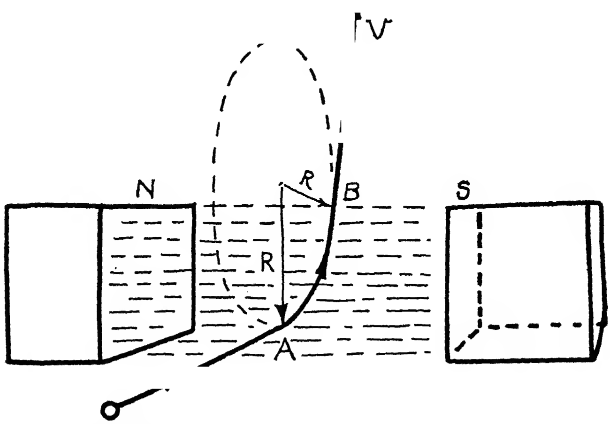


FIG. 72. The path of an electron which escapes from the magnetic field.

73 give the electron trajectories which an observer would see while facing the magnetic field. The dots indicate the field arrows. This principle of magnetic deflection of electron beams in the manner just described is used very generally in cathode ray tubes and television tubes for scanning.

Fig. 74 illustrates this application of magnetic deflection in its simplest form. Two solenoids, C_1 and C_2, are placed, as shown, at right angles to the axis of the cathode ray tube. This is a modern Braun tube in which an electron stream is focussed by

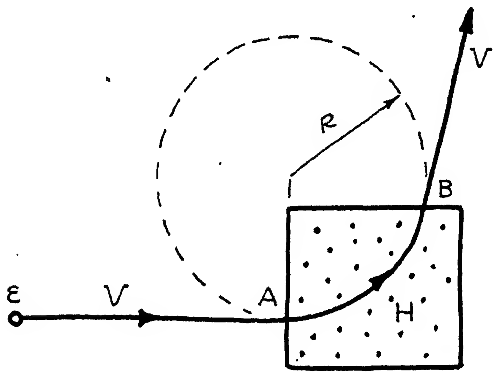


FIG. 73.

End-on view of Fig. 72. The dots represent sections of the magnetic lines of force.

electrostatic lenses so as to form a sharp spot of light on the luminescent screen S at the end of the tube. When the coils are activated by an alternating current, an oscillating magnetic field will be established along the axis A-B of the coils. Since the field, H, is proportional to the current flowing through the coils of the solenoids, it will reverse its direction continuously according to the alternations of the current, *i.e.*, H will be directed from A to B at one moment and from B to A 1/60th of a second later, if 60-cycle current is used. The effect of this alternating field, H, will be that the fluorescent spot on the screen, S, describes a vertical line flitting back and forth from P_1 to P_2.

Path of an Electron Moving in any Given Direction in a Magnetic Field: Vectors

Let us now investigate what takes place when the magnetic field and the electron beam do not happen to be at right angles to

each other as in the previous examples but are inclined to each other at a smaller angle.

For this purpose we need to say a few words about directed quantities in physics, which are called vector quantities. As examples we mention force, velocity, field strength; each of these quantities is determined by a numerical value and a direction. The velocity, or speed, of a moving object for instance, may be 50 miles per hour where 50 is the numerical quantity. In order to describe the movement fully, we must also indicate the direction in relation to some reference point. Thus we say, a car is

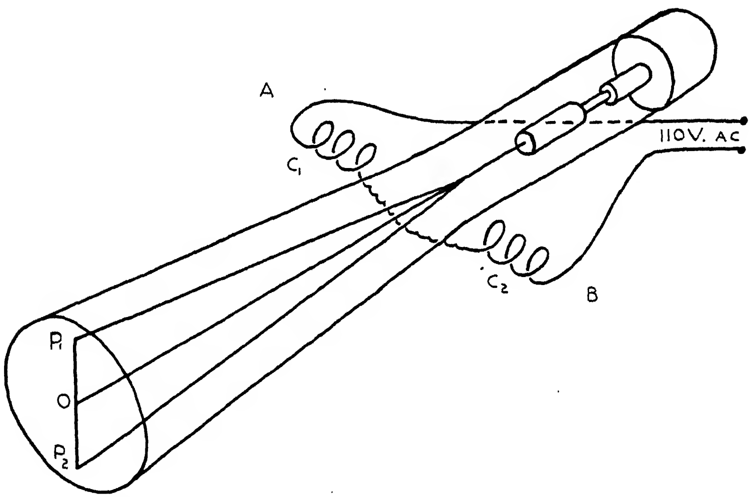


FIG. 74. The magnetic deflection of an electron stream due to an alternating field.

travelling north on a highway with a speed of 50 miles per hour. Everybody then knows exactly what the car is doing. Other quantities in physics are fully described by a mere number. They are called scalars; examples of scalar quantities are mass and temperature; the concept of direction is not associated with these terms.

Vector quantities are generally represented by arrows where the length of the arrow indicates the numerical value and the arrow head, the direction; the arrow is called a vector—merely a straight line having a definite length and a definite direction.

We have already made use of this symbolism in Fig. 59 (p. 120) where we describe the directional emission of electrons from a cathode.

The interesting feature of these vectors is that they can be split up graphically into various components or, what amounts to the same thing, several vectors can be added by simple rules in a diagram. Let us illustrate this rule for a force diagram.

In Fig. 75, a force F_1 of 3 units, acts upon a mass situated at O in the direction O-A for a certain time so that the mass is moved from O to A. A second force, F_2, of 4 units is then applied for the same time at A in the direction A-B so that the mass is moved to B. The same result could have been obtained by having a

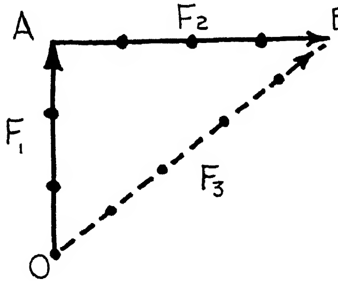


FIG. 75.

The sum of two vectors.

force, F_3, equal to 5 units act on the mass at O in the direction O-B for the same time. As the vector F_3 accomplishes the same effect as F_1 and F_2 both acting together, F_3 is said to be the sum of F_1 and F_2. We could also have applied F_2 first to get to C (Fig. 76a) and then F_1 to get from C to B and would have obtained the same result. By drawing the two arrows F_1, F_2 from the common origin O (Fig. 76b), we find that a rectangle results of which F_3 is a diagonal. If the two original forces act at a different

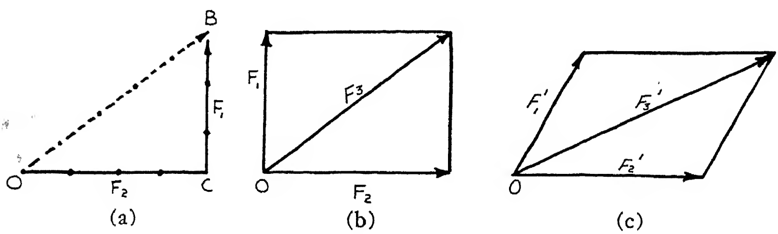


FIG. 76. The sum of two vectors.

angle with respect to each other, as in Fig. 76c, we obtain a parallelogram. In every case, the diagonal of the rectangle, or parallelogram, obtained in this graphical manner gives the magnitude and direction of the force F_3 which brings about the same final displacement as that produced, in steps, by F_1 and F_2. This is the well known parallelogram of forces probably familiar to the reader from his school days. Instead of forces, the arrows may represent velocities or other vector quantities.

By reversing the vector addition, we find that we can split up a vector into two components that add up to the original vector as long as the original vector forms the diagonal in a rectangle or parallelogram.

If the arrow marked v in Fig. 77a represents the magnitude and direction of a velocity with which a body moves from O, we will get to B by moving in two steps, v_1+v_2, as shown in Figs. 77b, and 77c.

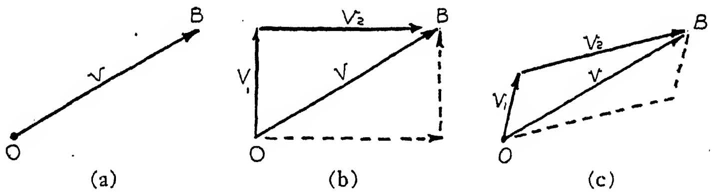


FIG. 77. A vector split into two components.

After this little excursion we now return to our main issue, *i.e.*, the influence of a magnetic field on moving electrons. In Fig. 78, we picture an electron moving with a velocity v in a uniform magnetic field H, so that the direction of v and that of H make an acute angle. We have no rule that would enable us to predict directly the trajectory which the electron will follow. But we do know the effect of H upon electrons moving parallel with, or at right angles to, H. Here we find the breaking up of a vector to be a welcome solution to our problem. We simply assume for the moment that the electron moves first with a velocity, v_1, at right angles to H and then with a velocity, v_2, in the direction of H and deduce what path will result when both v_1 and v_2 are possessed by the electron at the same time.

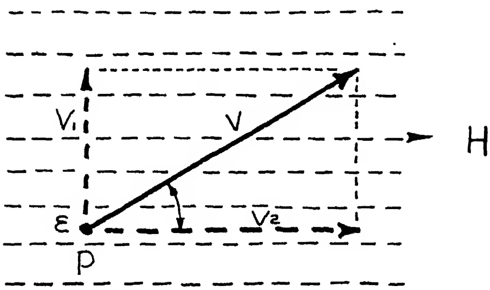


FIG. 78. The velocity of an electron in a magnetic field resolved into two new directions, one parallel to the field and the other at right angles to the field.

It is quite plain from the rules which we have stated above that the electron moving with velocity and direction v_1 in Fig. 78 describes a complete circle which is oriented at right angles to the direction of H and thus returns to P from where it started. We are also certain that the electron moving with v_2 in the direction of H is not affected by H and thus proceeds in a straight line. These two motions are represented in perspective in Fig. 79. Here a long solenoid is drawn which supplies the axial field, H. It has to be long in order to insure a uniform field in its central section where we shall study the electron movement drawn in Fig. 79, through the cut-away shown. We assume that the electron has the velocity v at P, on the solenoid axis O-O′, which

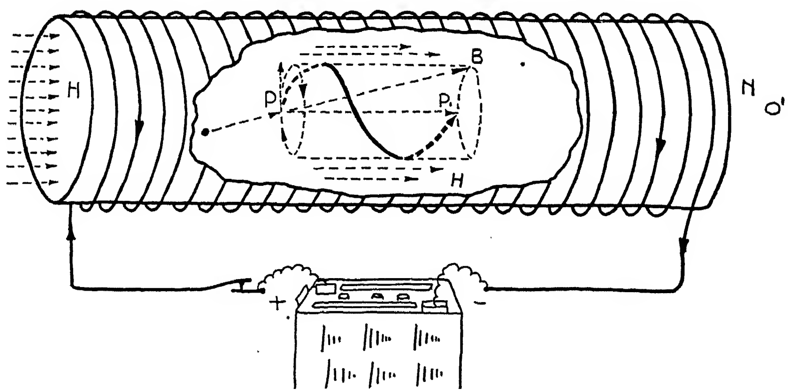


FIG. 79. The focussing action of a magnetic field on an electron diverging from the direction of the field.

of necessity will be the axis of a vacuum tube inserted into the solenoid. What happened to the electron before it arrived at P does not concern us here as long as we assume that it had a linear motion in the direction, PB. The moment it arrives at P the switch which actuates the solenoid is closed and produces the magnetic field instantly. Now we split v into v_1 and v_2 and ask what path the electron will follow when it has these two velocities simultaneously. The electron tends to describe the circle but at the same time it progresses with the uniform velocity, v_2, toward the right. The path thus becomes a helix which is wound about a fictitious cylinder touching the axis O-O′. Fig. 80 shows the view obtained by looking into the solenoid from the right. By

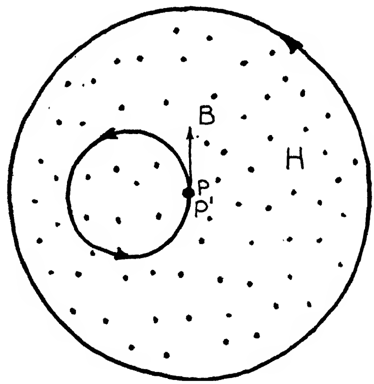


FIG. 80.
View of previous figure end-on.

the time the circle would have been completed the electron has arrived at P_1. In the absence of a magnetic field, the path of the electron would be along the straight line to B. We thus make the important observation that an electron which tended to move away from the axis (in a divergent beam) has been forced back onto the axis some distance away, because of the action of the coaxial magnetic field. This coaxial field, then, acts upon a divergent beam of electrons in the same way as a lens acts on a divergent beam of light, *i.e., it produces a focus.* Fig. 81 is the drawing of a space model which illustrates the trajectories of a number of electrons which originate at P with the same constant

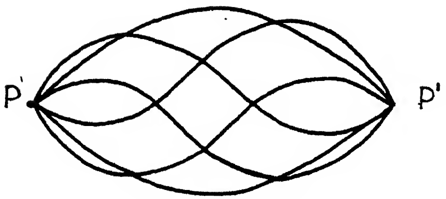


FIG. 81.
The helical paths of electrons from source P to focus P′.

velocity but in different directions with respect to the axis. They all converge toward P_1, the focus of the magnetic lens.

The lens property of the long solenoid was recognized at an early date. E. Wiechert in 1899 introduced the concentration coil, as it was called, in connection with the Braun tube and it proved to be a most useful device in the early days of cathode ray tube technique.

The Magnetic Lens

What we have described above at great length is really a very old story. The great advance made by Busch in 1926 was that he demonstrated and proved mathematically that a non-uniform magnetic field, in the direction of the tube axis, produced by a *short* solenoid also has the properties of a lens, and that the relation between focal length, magnetic field strength and electron velocity fitted well-known optical formulas.

The action of a short magnetic lens is illustrated diagrammatically in Fig. 82. A divergent electron beam originating at P on the electron optical axis of the system enters the non-uniform magnetic field produced by the short solenoid, S, and is thereby brought to a focus at P′. The short solenoid thus fully satisfies the optical conditions for the formation of an image of an object. A point in the object is reproduced as a point in the image. In optics, the object must be the source of light beams; in electron optics, the object at P must be the source of electron beams. In optics, the refractive medium is a glass lens; in electron optics, the refractive medium is a magnetic (or electric) field. In optics, the image is formed either on a screen or on the retina of the eye of the observer; in electron optics, the image is formed on a fluorescent screen or on a photographic plate.

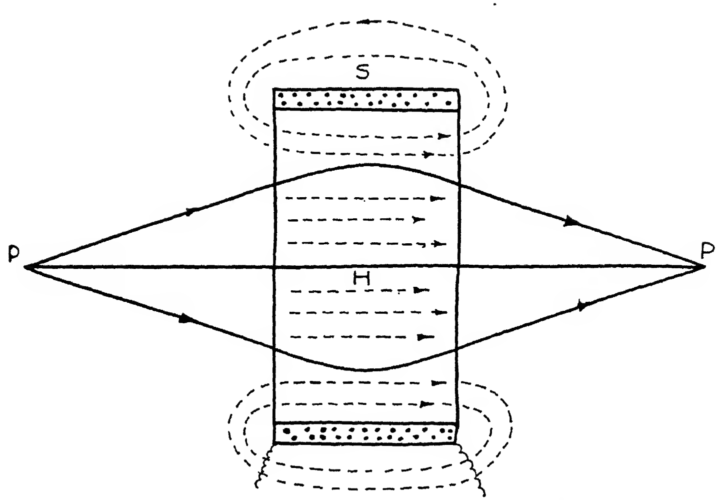


FIG. 82. The focussing action of the field of a short solenoid.

The equivalence of the action of a magnetic lens on an electron beam to that of a convergent glass lens on a light beam is further illustrated in Fig. 83, where the two corresponding systems of lenses are shown side by side. The focal points F and F_1 on each side of the lens are also indicated. The distances along the axis from the center of the lens to the object P, the image P′, the first and second focus F and F_1 respectively, are connected by a mathematical equation which applies in both cases.

The transformation of an object point into an image point has been demonstrated in the previous examples for the case that both these points were situated on the axis of the system. An object of physical dimensions must of necessity extend beyond the axis. Just as in light optics, the rule applies in electron optics, namely, that an image of objects can be formed as long as the image-forming rays do not diverge too far from the axis. Rays which travel near the axis are called *paraxial* rays. The means employed to insure that this condition is fulfilled are very similar for light rays and electron rays. It is the aperture stop familiar to everyone from the photographic camera. Fig. 84 illustrates the function of this device for limiting the beam angle. Thus in Fig. 84a, the light emitted by the object O at rather large angles

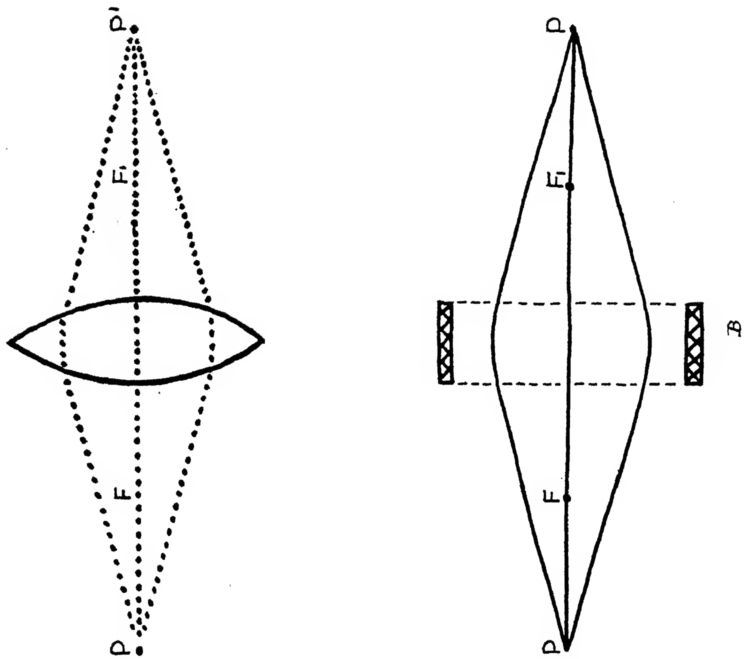


FIG. 83. Comparison of a convergent light (glass) lens and a convergent electron (magnetic field) lens.

with respect to the axis is stopped by the aperture disc A, so that it cannot reach the lens.

It has been mentioned in an earlier chapter that such wide-angle rays would cause severe distortions of the image if they were permitted to penetrate the marginal regions of the lens. This state of affairs also prevails in electron-optical lens systems. Fig. 84b shows the electron rays originating at an oxide cathode, C, which is again shielded by a guard ring, S. The anode, A, which accelerates the electrons serves, at the same time, as aperture stop and insures that only paraxial electrons reach the magnetic lens, L.

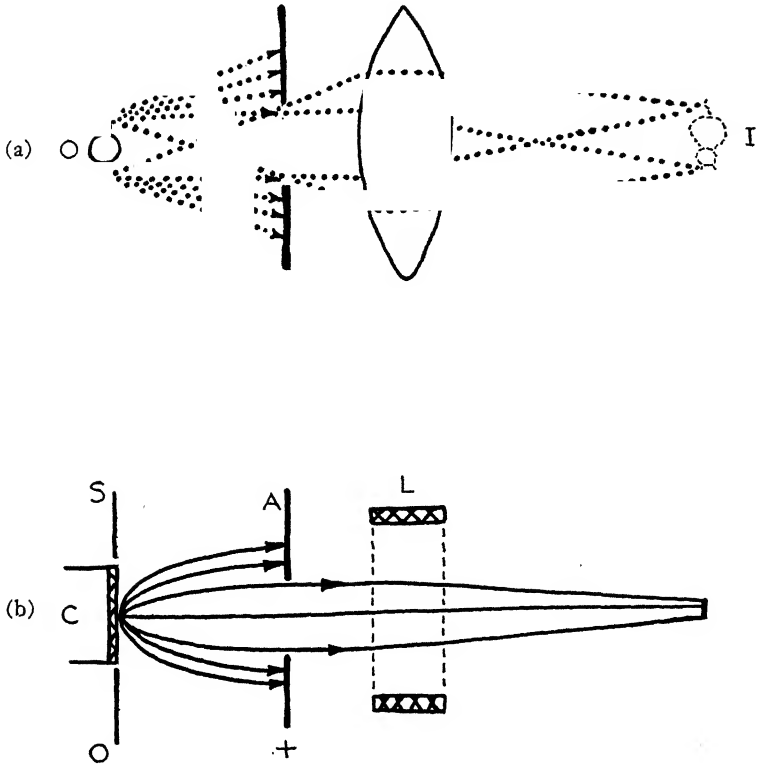


FIG. 84. Comparison of the use of aperture stops in light and electron lens systems.

Courtesy RCA Laboratories

Magnesium oxide smoke (×50,000).

Chapter 13

The History of the Electron Microscope

In the previous pages we have made ourselves familiar with the general principles and the outline of the theories that have led to the discussion of electron lenses. From now on, it will be our task to show how these lenses can be used to produce enlarged images of different types of objects in electron microscopes. As there exist two types of electron lenses, electrostatic and magnetic, we shall find that two types of electron microscopes are in use.

Before entering upon their description, it might be well to pause for a moment and review the steps in theory and experiment which resulted in this modern tool of research. This will give us an opportunity to record in chronological order the contributions which have been made from time to time by workers in different countries.

In optics, we find that lenses were used to form images of objects long before the dual nature of light was established. Even today, the designer of optical instruments leans heavily on the corpuscular concept in the sense in which it was associated with Newton, but has also to keep in mind that waves are involved and limit his simple geometrical theory. He is thus very old-fashioned and very modern at the same time. When he wishes to determine the path that a light beam will follow through a number of lenses, he proceeds to draw straight lines according to geometrical rules. We have used this technique in Chapter 2.

The field of application where this method gives satisfactory results is called *geometrical optics*. It covers the effects of reflection and refraction and lends itself particularly to ray tracing on mirrors and through prisms and lenses. The effects which

require the wave concept for their explanation form the domain of *physical optics*. Here we find the treatment of diffraction, interference and polarization. Since we realize that the corpuscular and the wave concepts cannot be segregated in principle when dealing with the nature of light, we may object to this classification of geometrical versus physical optics. After all, it is the same light which produces the effects in both domains. Nevertheless, this terminology has been carried through the textbooks and may have its use at times.

We should then, likewise, speak of geometrical electron optics in cases where the corpuscular concept of the electron is uppermost in our minds, and of physical electron optics where we stress the wave nature of the electron, while being fully aware of the dual nature of the electron at all times. This practice has indeed been followed.

When we trace the trajectory of an electron beam through a potential field we use a set of geometrical rules—different of course from that used in optics—and obtain a first approximation to the actual path. For this reason, the treatment of electron lenses and the associated electron paths form the subject of *Geometrical Electron Optics.* Davison and Germer's experiments on the diffraction of electrons in a crystal lattice and the discussion of the resolving power of the electron microscope where the wave nature of the electron plays the decisive role belong to the field of *Physical Electron Optics.* At times it may be doubtful which department may solve a certain problem the more easily. The reader should not worry about this; it is not very important.

We may now state that geometrical electron optics was developed as a result of—and not before—the full understanding of the dual nature of the electron. This need not have been so.

Long before the electron was discovered, Sir William Rowan Hamilton had clearly shown (in 1830) that a striking analogy exists between the law that governs the movement of a corpuscle through a given field of force and the law which governs the trajectory of a light beam through a series of media of different refractive indices.

In the first case, the mass will move from a point P_1 to a point P_2 under the influence of the existing mechanical forces in such a manner that the product of the momentum, mv, multiplied by the length of a small element of the path, ds, will give the smallest possible value when added up over the total length of the actual path from P_1 to P_2. Readers who have been exposed to a mathematical treatment of mechanics will recognize in this rule the famous principle of Maupertuis, which was formulated in the middle of the eighteenth century and which may be given the following mathematical form:

$$\int_{P_1}^{P} v\,ds = \text{a minimum}$$

In the case of light, Fermat had shown, in 1667, that a beam of light will follow a path from a point P_1 to a point P_2 through various media of different refractive indices in such a manner that it takes the least possible time. This principle can be expressed mathematically in the following form:

$$\int_{P_1}^{P_2} n\,ds = \text{a minimum}$$

where n is the refraction index of the medium at any point.

Putting these equations side by side reveals to us, who are by now supposedly well educated by the literature of the preceding chapters, the fundamental similarity between light and matter. We find as we have stated before in Chapter 11 that the particle velocity, v, takes the place of the refractive index, n, in the optical laws.

When the electron was discovered and recognized as a particle of a definite mass, m, and electron experiments had revealed striking similarities to optical experiments (1900), the way was open to develop geometrical electron optics on the basis of Hamilton's analogy between the basic mechanical and optical principles. By the time the dual nature of light was generally accepted by physicists, about 1910, there was really no excuse for not

suspecting a like duality of the electron. Evidently, such a sympathetic approach to the study of nature, which would have been natural to the Greeks, was not cultivated at the time.

So the years went by, existing theories were carefully refined and much experimental work done with the aim of revealing the structure of the atom and the role which electrons played in it. Many times, when electron beams were caused to interact with atoms, whether they were in the form of a gas or in a solid array at the surface of a crystal, the observed effects were not in agreement with the conventional view of the corpuscular nature of the electron.

We then witness, in 1923, the announcement of the dual nature of the electron by Louis de Broglie. There was as yet no experimental proof; a great vision had been crystallized into mathematical theory and stood as a challenge to the experimental physicist. Strangely enough this challenge was not accepted immediately. It was necessary for W. Elsasser (1925) to call attention to the fact that the demonstration of electron diffraction would establish beyond a doubt the postulated wave associated with the electron. C. J. Davisson, of the Bell Telephone Laboratories in New York, began to search for this diffraction effect systematically in 1926 and was joined in this work by L. H. Germer a little later. In 1927 their efforts were crowned with success. The experimental data obtained from the diffraction of fast electron beams at the surface of a nickel single crystal confirmed the de Broglie relation for the electron wave length,

$$\lambda = \frac{h}{mv}$$

Similar effects were observed by G. P. Thomson and his co-workers at about the same time. In 1926-27 H. Busch established the theory of the electron lens represented by axially symmetric magnetic or electric fields, and thus became the founder of geometrical electron optics.

In 1931 C. J. Davisson and C. J. Calbick performed experiments on the lens properties of apertures and published the formula

for the focal length of a circular and a slit aperture. Thus the electrostatic electronic lens made its appearance.

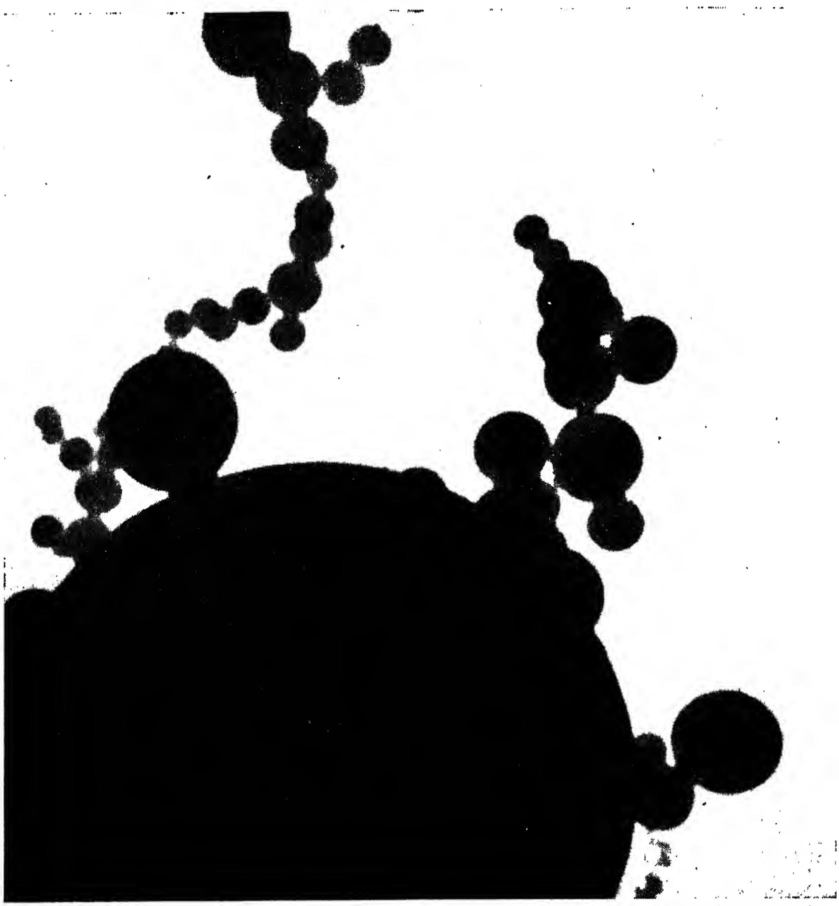
During the ensuing years, the subject of geometrical electron optics was most actively developed in Germany and a large number of publications in scientific journals followed one another in rapid succession.

In 1932 both types of electron microscopes were described and the first images were shown. E. Brüche and H. Johannson produced the first electron images of an oxide cathode with an aperture lens system utilizing 300-volt electron beams. M. Knoll and E. Ruska developed the first magnetic electron microscope and produced images of a cold cathode with short magnetic lenses, utilizing 60,000-volt electron beams. The theory of electron lenses was largely developed by J. Picht, O. Scherzer and W. Henneberg.

The amount of work covered by this school of investigators within the short time of a few years may be appraised by the fact that in the latter part of 1934 the first book entitled "Geometric Electron Optics" (in German) was published by E. Brüche and O. Scherzer. It consisted of some 320 pages of scientific and technical matter and has become a classic to workers in this field. The years from 1934 to the present have not produced any fundamentally new principles in electron optics, if such were at all possible, but have brought about a continuous improvement of electron-optical instruments and a better understanding of their mode of operation. Some of these features will be dealt with in the ensuing chapters. The work on the construction of an electron microscope at the McLennan Laboratory, University of Toronto, has already been recorded in the preface.

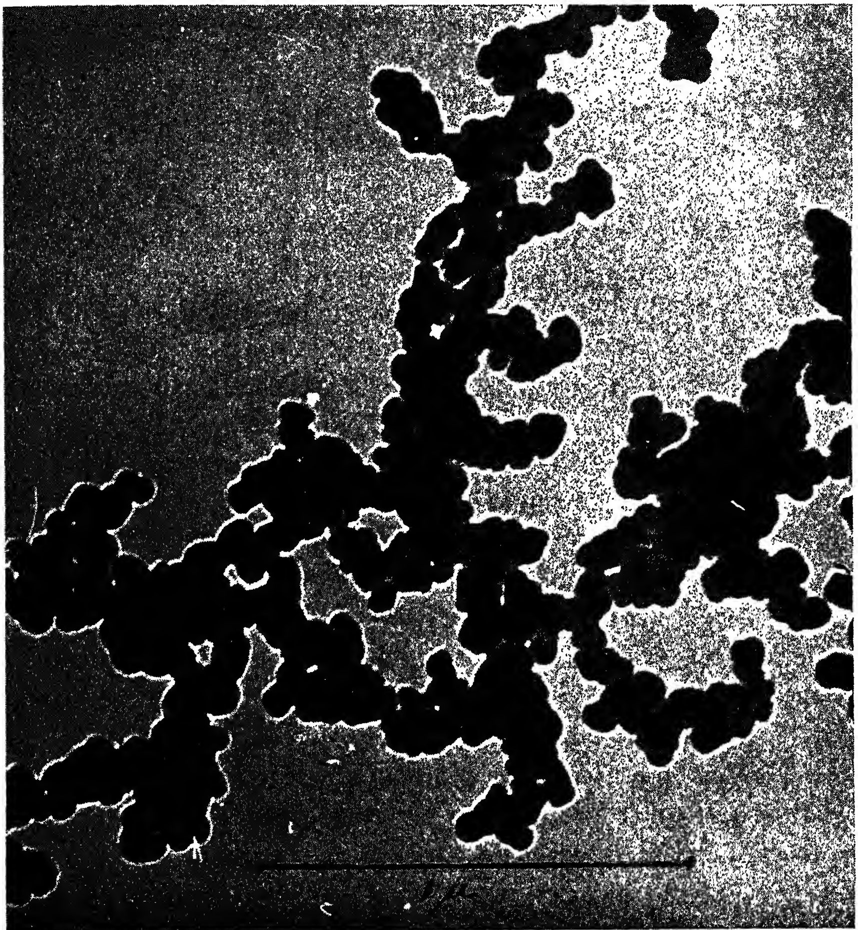
This rapid development carried on in Toronto is only one example of the concentrated efforts that were directed towards the improvement of electron microscopes in many other places—so much so that it is now difficult to state how many of these instruments exist.

We shall now proceed to give in the following chapters a description of the structural features of the electrostatic and magnetic electron microscope as far as it is essential to the understanding of their operation. Very many details must of necessity be omitted in a general text of this type.



Courtesy American Cyanamid Co.

Aluminum oxide smoke (×30,000).



Courtesy RCA Laboratories

Carbon black particles (×60,000).

Chapter 14

The Electrostatic Electron Microscope

Object and Image

In order to form images of objects with the aid of electron beams, one essential condition must be satisfied. Electrons which leave a point on the object at small angles with respect to the normal to the object plane must be reunited at a point in the image plane. Object and image plane are parallel to each other and separated by a given distance along the electron optical axis of the system. Object points in the object plane are confined to the close proximity of the axis, whereas image points in the image plane are located within the circle of sharp image formation which will have a diameter of several inches depending on the magnification of the system. These statements are illustrated in Fig. 85 in a simplified manner. Five representative points A, B, C, D, E are chosen in the object plane (O, P) and form the apex of five cones of divergent electron rays. In the plane indicated by L an electron lens, or a combination of several such lenses which in effect can be replaced by an ideal thin lens, is located. The function of this lens is twofold. The cone of electron rays which diverges from the object point must be inverted so as to converge to a point in the image plane. The lens plane thus becomes the base of a series of double cones. In order to bring about this result the forces acting upon the electrons in any one of the various beams must be directly proportional to the radial distance of an electron from the axis of any one beam to which it belongs. In this manner, the electrons which tend to diverge the most undergo the greatest constricting force. At the

same time, the electron lens will impart to the beams a deflection which will make the distances between the image points, A′B′C′D′E′, respectively, greater than the corresponding distances between the points, A B C D E, in the object plane. Thus a magnification of the object results which is given by the ratio of the distance of an image point from the optical axis to the distance of the corresponding object point from the axis. This ratio M, the magnification, must be constant for all image points in order to obtain a linear representation of the object in the image plane.

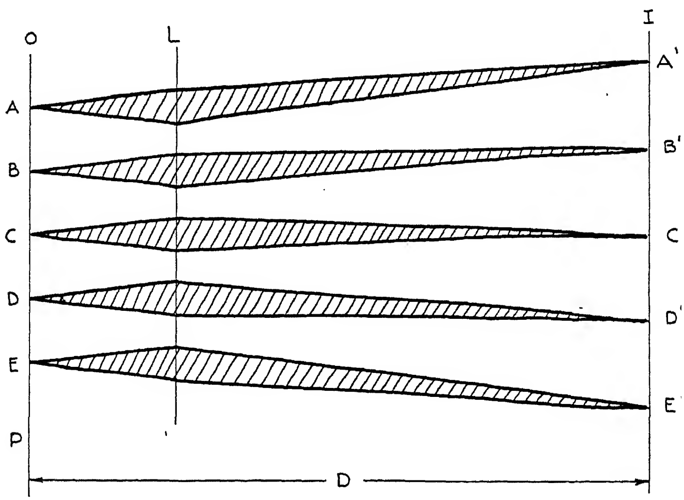


FIG. 85a. Object points and image points with an electron lens. Equivalent to a negative lens.

Figs. 85a and 85b illustrate how the electron beams may either spread out from the lens plane or cross each other in an intermediate plane, F, near the axis. (It may be remarked here, that this socalled "cross-over" is of particular interest in the theory of cathode ray tubes since it represents a virtual cathode of high current density. A limiting aperture is placed in this plane, F, and an image is formed of it on the fluorescent screen with as low a magnification as possible so as to obtain an intense and very small pin point of light on the screen.)

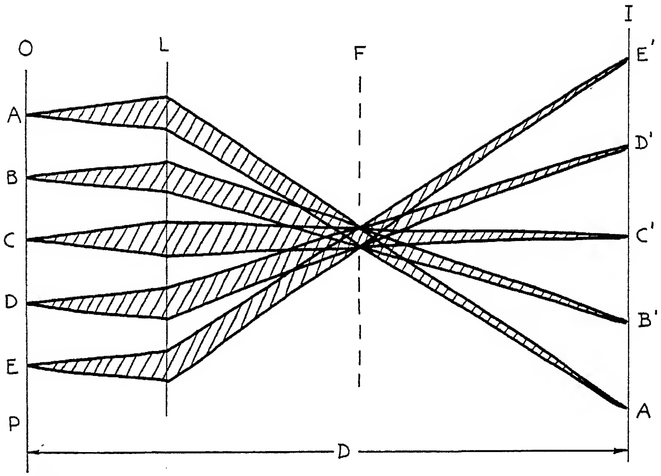


FIG. 85b. Object points and image points with an electron lens. Equivalent to a positive lens.

An Electrostatic Negative Lens

In the first case (Fig. 85a) the lens action would be equivalent to a negative lens in optics (see Fig. 12b) and could be brought about by an aperture lens in front of a cathode followed by a field free space to the right (Fig. 86). This arrangement represents the only negative electron lens known. All others are positive in their action (see Fig. 12a). Since the lens field in Fig. 86

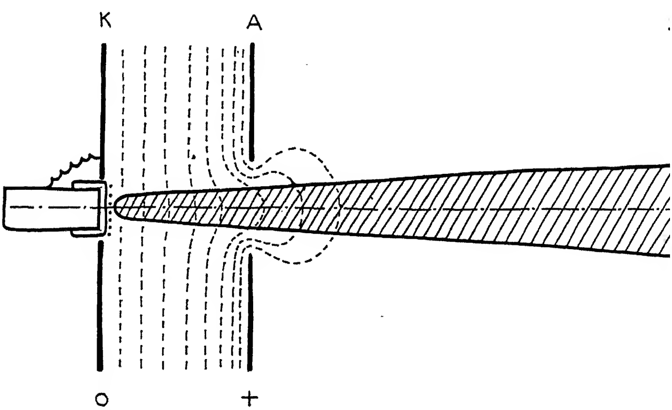


FIG. 86. The theoretical arrangement for a negative electron lens.

extends to the cathode, our new object, we have to classify this arrangement as an immersion lens. Unfortunately this system has no practical value because the voltage that can be applied in close proximity to the cathode is limited to values of the order of 50 volts and electrons of such low velocity barely excite fluorescence on a luminescent screen. We find then, in most electron image-forming systems, that the beams cross the optical axis and that the position of the image is inverted with respect to the object, as shown in Fig. 85b.

Electrostatic Positive Lenses: The Three Electrode Lens

The first practical electron lens system which lent itself to the construction of an electrostatic electron microscope was de-

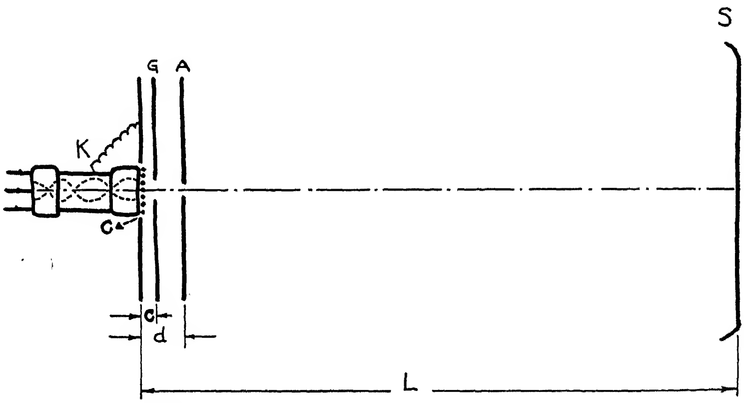


FIG. 87. Brüche and Johannson electrostatic electron microscope.

Diameter C =	5 mm.	L = 240	mm
Aperture G =	1.2 mm.	V_G = 140	volt
Aperture A =	1.0 mm.	V_A = 750	volt
c =	0.5 mm.	V_S = 750	volt
d =	1.0 mm.		

scribed by Brüche and Johannson in 1932. It is called "the immersion objective lens." We shall here briefly describe the structural features of this lens without going into a discussion of its theory which is rather involved. For purposes of brevity we shall use the letters I.O. as an abbreviation for immersion objective.

In its original form, the I.O. consisted of two aperture disks located in front of the cathode, as shown in Fig. 87. The indi-

rectly heated cathode, C, of the oxide-coated variety, is surrounded by a guard ring at cathode potential as we have described on other occasions in this text (see Fig. 58). In front of it, at a distance, c, from the cathode surface, a first disk of molybdenum sheet, 0.2 mm thick, is mounted with aperture G 1.2 mm in diameter. In conformity with radio-tube practice this disk element is called the grid, G. At a distance, $d=1$ mm, from G a second molybdenum disk, A, with an aperture 1 mm in diameter, is mounted; this serves as anode. The fluorescent screen, S, which is maintained at anode potential, is mounted at a distance, $L=240$ mm, from the cathode surface. It goes without saying that the various electrodes must be aligned most accurately with respect to

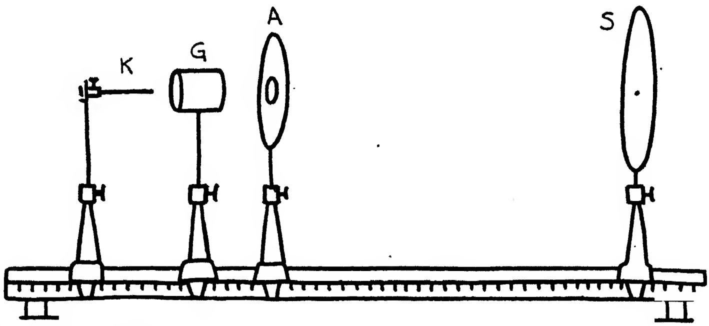


FIG. 88a. Arrangement of part of the electron microscope on an optical bench built inside a vacuum tube. (Fig. 87)

the axis and with respect to one another, since the slightest deviation from axial symmetry will produce image distortions. The aperture edges must not only be perfectly round but also free from burrs and carefully polished. In practice, this accuracy of the mount is achieved by the use of ceramic spacers between electrodes. In an experimental setup such as that of Johannson, the elements were mounted on an optical bench along which they could be moved *in vacuo* by means of external magnets. Provision was also made for varying the distance, c, between cathode and grid aperture, G, by mounting the cathode on a system of bellows. Fig. 88 a, b, c, shows (a) the elements on the optical bench sup-

port, (b) the cathode bellows, and (c) cathode images* obtained with the I.O. By adjusting the voltages on the electrodes C, G, A, it was found that a sharp image was obtained for $V_G = 140$ volts and $V_A = 750$ volts corresponding to the dimensions given in the legend of Fig. 87.

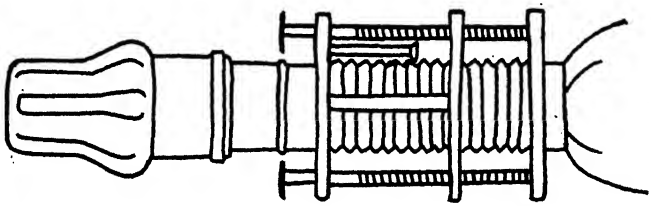


FIG. 88b. Arrangement of bellows for adjustment of the position of the cathode in a three-electrode lens.

The Magnification of a Three-Electrode Lens

Fig. 89a shows how the grid potential V_G must be adjusted in order to maintain a sharp image when c is varied. Fig. 89b gives the resulting effect on the magnification, M.

It turns out that, for any given value of c, a sharp image is obtained with a definite value of the ratio V_G/V_A, and the magnification, M, depends only on this ratio, irrespective of the mag-

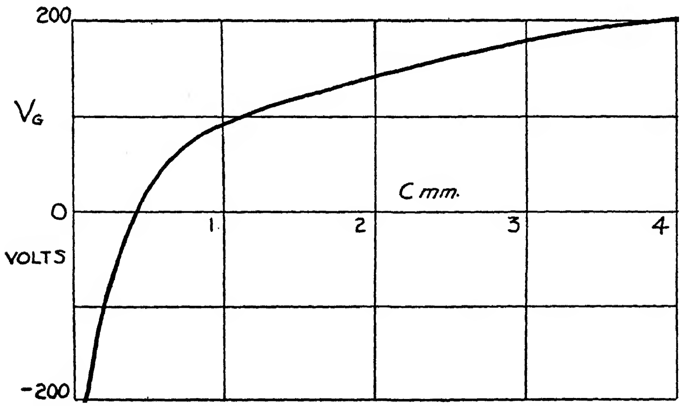


FIG. 89a. Graph showing the value of the grid potential, V_G, necessary to give a sharp image in the three-electrode lens as the distance, c, between the cathode and the grid is changed.

* For Fig. 88c, see C, E, G, H and I on plate, p. 176.

nitudes of V_G and V_A which satisfy this ratio. This is a very general law for electrostatic electron lenses; their power depends always on a potential ratio. In Fig. 90 the magnification, M, the diameter of the sharply reproduced image B_0, the diameter of the

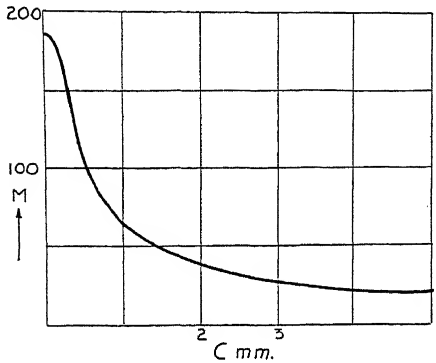


FIG. 89b.

Variation of the magnification, M, of a three-electrode lens with the distance, c, between cathode and grid.

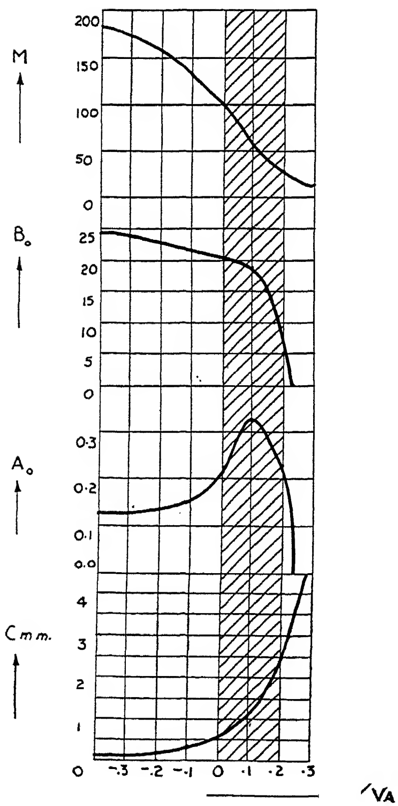


FIG. 90.

How (1) the magnification (M), (2) the diameter of the sharp image (B_0), (3) the diameter of the region on the cathode ($A_0 = B_0/M$) which is reproduced in the image, and (4) the distance between cathode and grid (c), severally vary with the ratio of grid potential (V_G) to anode potential (V_A), for the three-electrode lens.

sharply reproduced region on the cathode $A_0 = B_0/M$ and the distance, c, are plotted against the ratio V_G/V_A as abscissa.

If one aims to reproduce sharply as large an area of the cathode as possible, the I.O. must be operated at a value $V_G/V_A = 0.1$, for which A_0 is a maximum in Fig. 90, and c must be made equal to 1.0 mm. The magnification then is about fifty-fold. We also learn from the characteristic curve that M reaches values as high as 200 if we are satisfied to reproduce smaller areas of the cathode. The cathode must then be very close to the grid aperture and the latter will assume negative potentials according to Fig. 89a. This

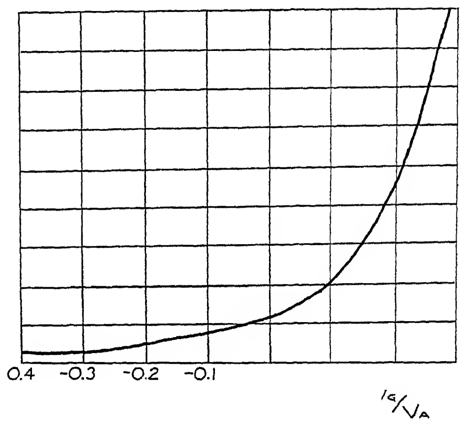


FIG. 91.

How the reduced focal length of the three-electrode lens, F_1, varies with the ratio V_G/V_A.

is in accordance with general microscope technique, where large magnifications are obtained with an objective lens of very short focus which demands that the object must be placed almost in contact with the lens.

Fig. 91 gives a plot of the reduced focal length of the I.O., which takes into account the length of the system. For the optimum value $V_G/V_A = 0.1$, a focal length of 3 mm. results. This curve also demonstrates the important fact that the focal length of an electron lens can be varied at will by changing the ratio of the applied voltages in the case of an electrostatic lens. This permits of considerable flexibility of the electron-optical system and allows adjustments of the power of the lens over a fairly wide

range. A glass lens has only one fixed value for its focal length, which is determined by its geometrical shape and the type of glass used. Again, the refractive index of the glass does not exceed 1.66, while for electron lenses the quantity corresponding to the refractive index, *i.e.*, $\sqrt{V_2/V_1}$, may exceed 10.

The Four-Electrode Lens and its Magnification

A modified I.O. of superior quality was described by Johannson in 1934. It differs from the system just described by having an additional electrode which makes it a four-electrode system.

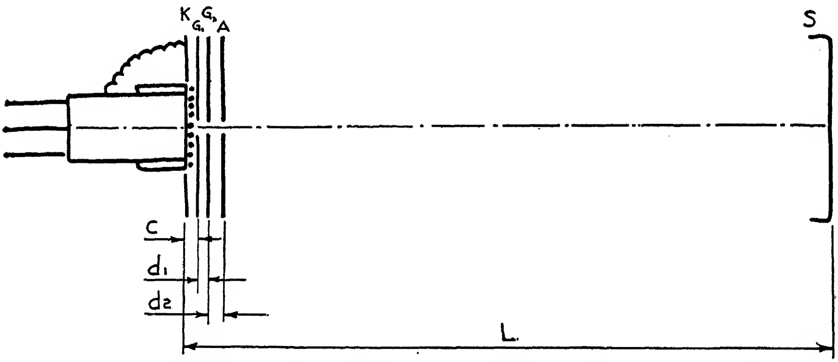


FIG. 92. Johannson's four-electrode electrostatic electron microscope.

$D_c =$ 5 mm.		$c =$ 0.05-1.5 mm.	
$D_{G1} =$ 1.1 mm.		$d_1 =$ 0.48 mm.	
$D_{G2} =$ 1.0 mm.		$d_2 =$ 0.72 mm.	
$D_A =$ 1.0 mm.		$L = 240$ mm.	

Fig. 92 gives the diagrammatic view. The performance of the four-electrode I.O. is superior to that of the three-electrode system by about 50 per cent, as measured by a figure of merit B/L; B is the diameter of the sharply reproduced image. It makes possible not only a larger useful magnification, by 25 per cent, but also permits of the variation of M over a wider range. The focal length is correspondingly reduced to the order of 1 mm. The following figures give the performance characteristics of this system.

The potentials to be applied to the various electrodes, for a given cathode-to-G_1 distance c, are readily derived from the

graphs in Figure 93, where the ratios V_{G_1}/V_A and V_{G_2}/V_A are plotted. Thus, if we choose $V_A = 1000$ volts, we may decide to get a large magnification, M, of a small circular region of

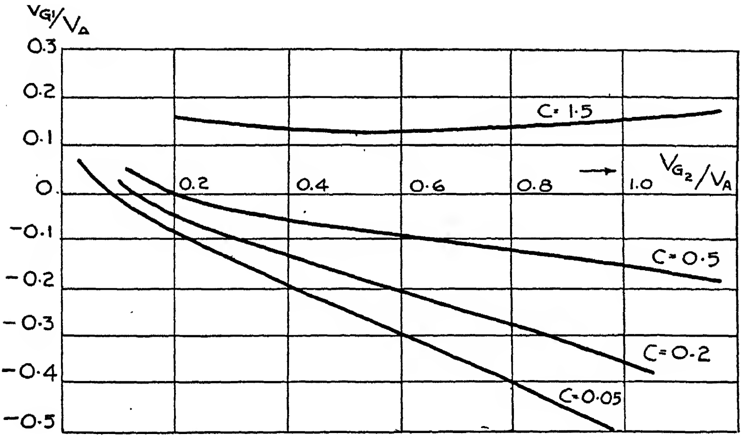


FIG. 93.

diameter, A, of the cathode. Figure 94 then indicates that we should make $V_{G_2}/V_A = 1.0$ and give as small a value as practical to c, let us say 0.2 mm. (or 8 mils.). The potential of G_2 is then equal to that of the anode A, *i.e.*, 1000 volts. The intersection of the curve marked $c = 0.2$ (in Fig. 93), with the ordinate at

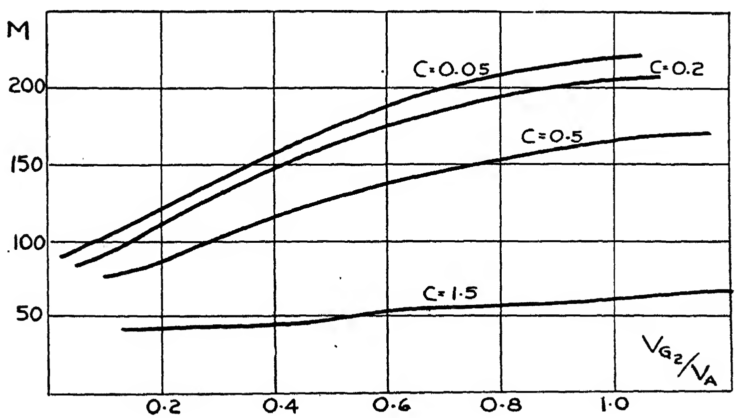


FIG. 94.

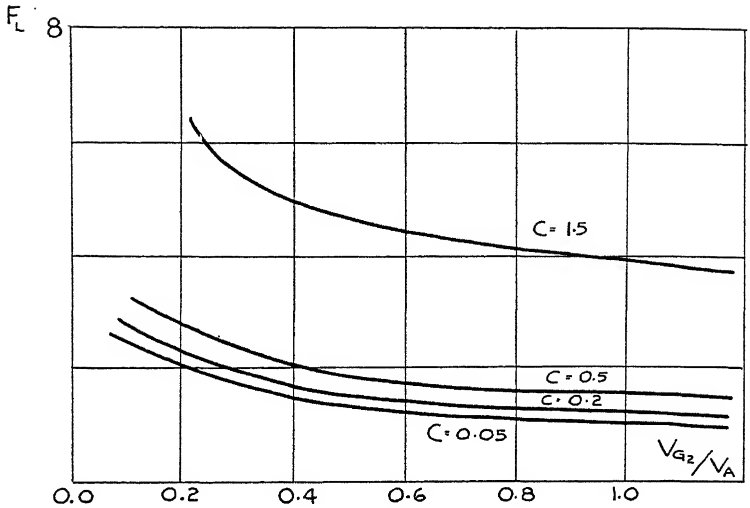


FIG. 95.

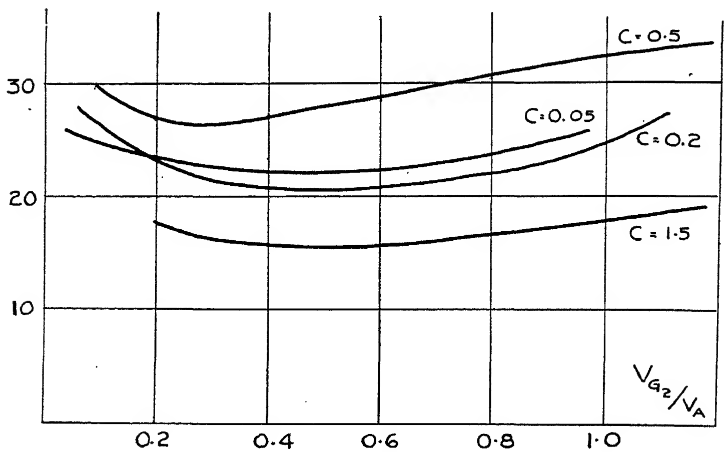


FIG. 96. Variation in B.

at $V_{G_2}/V_A = 1$ locates a value $V_{G_1}/V_A = -0.35$, which gives $V_{G_1} = -0.35\ V_A$. The potential of G_1 is negative and, in magnitude, equal to 35 per cent of V_A, or -350 volts. This will give us a magnification of about 220 for the dimensions of the system as stated in the legend of Fig. 92. The diameter, A, of the region on the cathode producing the image is about 0.12 mm. as obtained

from Fig. 97, *i.e.*, less than 5 mils, which is about the thickness of a sheet of common note paper. Much larger magnification can be obtained with the electrostatic electron microscope if all the dimensions of the I.O. are reduced by a constant factor. $M = \times 4{,}000$ has thus been reported by Johannson. Since the size of the aperture must also be correspondingly reduced we would obtain aperture diameters of the order of 0.1 mm. (4 mils) when using a reduction factor 10 to get $M = \times 2{,}500$. This will necessarily weaken the intensity of the beam and reduce the

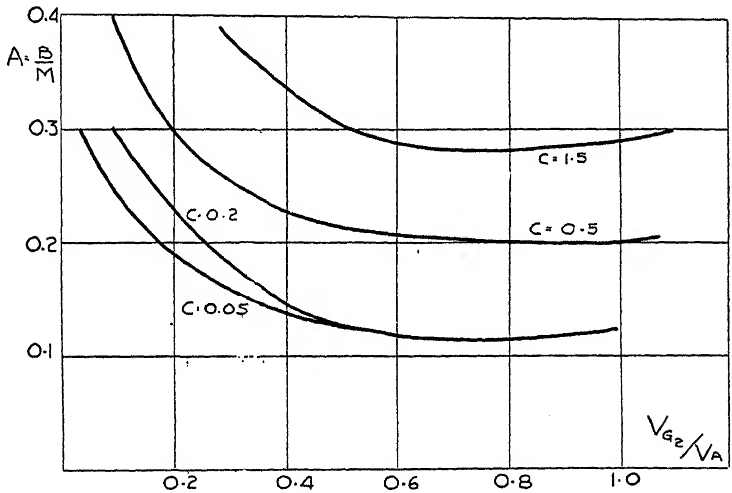


FIG. 97.

brightness of the image obtained on the screen. For this reason the instrument is generally operated to give a magnification of the order of a few hundred times.

The Resolving Power of the Electrostatic Lens

When we come to consider the resolving power of the I.O., we must remember that it depends on the electron wave length in the object plane. The initial velocities of electrons emitted from an oxide cathode are of the order of 0.1 volt. Introducing this value into de Broglie's formula for the electron wave length (see p. 95),

$$\lambda = \frac{12.24}{\sqrt{V}} \text{ A.U. } (\text{A.U.} = 10^{-8} \text{ cm.})$$

we obtain:

$$\lambda = \frac{12.24}{\sqrt{0.1}} \text{ A.U.} = 38.92 \text{ A.U. or } \frac{38.92}{100{,}000{,}000} \text{ of a cm.}$$

This value for λ must then be substituted in the formula given on page 38 so that we obtain for the distance, d, of two object points which can just be separated in the image: (0.61 replaces ½).

$$d = \frac{0.61\lambda}{\text{N.A.}} = \frac{0.61 \times 38.92 \times 10^{-8}}{\text{N.A.}} \text{ cm.}$$
$$= \frac{23.70}{\text{N.A.}} \times 10^{-8} \text{ cm.}$$

The numerical aperature N.A., ($= n \sin A$), as defined on page 38, is much smaller in electron-optical microscopes than it is in ordinary optical microscopes where values N.A. = 1.5 can be realized only with an oil immersion objective.

In electron-optical systems severe limitations are imposed on the size of the aperture diameter in order to keep lens aberrations to a minimum. These aberrations are caused by rays divergent from the optical axis beyond a permissible amount, as we have learned above. To prevent such excessive divergence is the purpose of a very small aperture diameter which leads to a reduced value of sin A. We have also found in Chapter 11 (page 116), that the electron velocity corresponds to the refractive index n in optical systems. Since we are dealing with very slow electrons in the I.O. this value likewise becomes very small, so that $n \sin A$ is smaller by three orders of magnitude in comparison to values common in ordinary optics; *i.e.*, N.A. = 0.001. If we introduce this value in the last equation we obtain

$$d = \frac{23.7}{.001} \times 10^{-8} = 23.7 \times 1000 \times 10^{-8}$$
$$= 23.7 \times 10^{-5} = 23{,}700 \text{ A.U. or 237 millionths of a cm.}$$

Actually, Brüche and Knecht determined experimentally that a resolution of 15,000 A.U. had been reached with their instrument (1934).

There are other factors which operate against obtaining high resolving power with the I.O. when it is used for the investigation of self-emitting objects, in particular of oxide-coated cathodes. The granular structure of the cathode surface will distort the first refracting surface of the electron lens, *i.e.*, the equipotential area which matches the surface of the cathode. A local negative space charge in front of the cathode produced by particularly active spots on the cathode may act in a similar manner. For refined investigations, it is then necessary to provide as smooth a surface as possible and to operate it at a relatively low temperature, so that space charge distortions are avoided.

In spite of these restrictions, the service of the electrostatic electron microscope in manifold researches has been invaluable. We shall give a brief survey in the following chapter of the various applications to which it has lent itself.

The Possibilities of Using Bright Emitter Cathodes

We have thus far considered only the case of an oxide-coated cathode being used as an electron source in conjunction with the I.O. Nothing stands in the way of using a bright emitter such as a tungsten or molybdenum ribbon as the object. Due to the elevated temperature at which these materials have to be operated in order to yield emission currents comparable to those obtained from dull emitters, a certain amount of direct light will reach the central area of the image and produce a background illumination. This need not be serious since the apertures are small in any case. The metallic surfaces of these cathodes can be highly polished so that field distortions are greatly reduced at the surface. In addition to this, higher voltages may safely be applied so that the resolving power will be increased considerably.

Photoelectric Surfaces as Cathodes and other Variations

Photoelectric surfaces can also be used as cathodes in the I.O. Provision must then be made that a light beam reaches the photo-cathode from the side, or from the rear when a transparent photo-cathode is employed. This latter principle is the basis of

Zworykin's image tube which has found many useful applications. Since the emission yield from photo-cathodes is comparatively small, it is necessary to apply voltages of the order of 10,000 volts to the lens anode in order to obtain images of sufficient brightness on the luminescent screen.

The cathode may be the source of secondary electrons produced by the impact of primary electrons either under angle from the side or from the rear in the case of a thin metal film.

Finally, high-speed electrons may be shot through a thin film, a technique which was explored experimentally by R. Behne in 1936. He was thus able to obtain images of aluminum foils 0.6μ thick with a magnification nearly 100 times, when 500-volt electrons were shot onto the foil from the rear and then accelerated by 3,000 volts and more on the anode of a three-element, Johannson Immersion Objective.

According to recent news reports the electrostatic electron microscope has been further developed in Germany by the Allgemeine Elektrizitäts Gessellschaft, Berlin, a company with which Brüche, Scherzer, Henneberg, Johannson, Behne, Recknagel, and Mahl are associated. This new instrument utilizes high-speed electrons and thus overcomes one of the main limitations which we have found to operate against the realization of a high resolving power.

Chapter 15

Applications of the Electrostatic Electron Microscope

Oxide-Coated Cathodes

In this chapter we shall confine ourselves to a review of investigations on self-emitting objects (cathodes) carried out with the electrostatic electron microscope. As this field of application is so large that a whole book could be written about it alone, we must be satisfied with a number of typical examples taken from the published accounts of many contributors in various countries.

Keeping the limited resolving power of the electrostatic electron microscope in mind, one might well ask what advantage is to be gained from the use of this instrument, since it can evidently not reveal any greater detail than can be obtained with an optical microscope. The answer is that electron emitters must of necessity operate in a vacuum, and observation by means of an optical microscope with short focus objectives is thus impossible. To be able to use the emitted electrons directly as the image-forming rays, and thus to obtain information about the emitting surface under operating conditions, is then of great value.

The study of oxide-coated cathodes was the first application to which the electrostatic electron microscope lent itself. Thermionic emission from such cathodes has been used technically for over thirty-five years, but only recently has the theory of the mechanism of emission been clarified.

Since the cathode surface should be uniform, the coarseness of the grain of the oxide layer (a mixture of barium oxide, stron-

tium oxide and calcium oxide) has a disturbing effect on the field pattern in front of the cathode. It is consequently desirable to use a very fine grain size for the coating in order to improve the quality of the images. Such fine grain coatings must be applied in such a way as to produce a uniform thickness over the cathode area. This is done by a kind of electroplating process called *electrophoresis* (or *cataphoresis*). The powder particles are so fine that when they are dispersed in a liquid they have the characteristics of a colloidal solution, the particles of which have diameters ranging from 2×10^{-5} to 10^{-7} cm. Such small particles cannot be seen with an ordinary optical microscope using visible light. The term *colloidal solution* is defined by this range of particle size. Colloidal particles have an electrical charge, positive or negative, as the case may be; they will migrate under the influence of an electrical field and can consequently be deposited on one of the electrodes immersed in the colloidal solution. Cathode surfaces and very fine hairlike filaments, such as are used in battery-operated radio tubes, are coated in this manner.

It turns out that electrophoretically coated cathodes have several technical advantages. As M. Benjamin and others have established by careful research, "the emission from an oxide cathode increases as the particle size of the oxide decreases. It is shown that the smaller particles give a more uniform emitting surface and increase the ratio of emitting to non-emitting areas." This result was obtained with the aid of the electron microscope. Von Buzagh has indicated that an optimum value of particle size exists which gives the greatest emission yield just below the upper limit of colloidal size.

Mechanism of Electron Emission from Oxide-coated Cathodes

It has been established by innumerable experiments and investigations during the past few decades that the copious emission of electrons from an oxide-coated cathode is due to the existence of a very thin film of metallic barium on the surface of the oxide layer.

The oxide layer itself is produced by a chemical reaction that takes place when the cathode is first heated *in vacuo*. It is probably the most general practice in America to spray onto the cathode surface a mixture of barium carbonate ($BaCO_3$) and strontium carbonate ($SrCO_3$), to which calcium carbonate ($CaCO_3$) is added in some cases. These carbonates are prepared in the form of a suspension in amylacetate, with some nitroglycerin as binder. Mixtures of this type are commercially available from chemical supply houses. The cathode material consists of pure nickel or nickel with a fraction of a per cent of titanium or aluminum added. During the process of exhausting the tube the cathode is heated in steps up to a temperature of about 1350°K. The alkaline-earth carbonates (or nitrates, or hydroxides as the case may be, depending on the compound that was originally sprayed onto the cathode) break down at about 1100° K into oxides and carbon dioxide (or nitrous oxide or hydrogen in the case of the nitrates and hydroxides, respectively). The gas is pumped away, leaving the oxide layer on the cathode. In the presence of small amounts of reducing agents such as carbon, titanium or aluminum, about 0.5 per cent of the oxide molecules are reduced to free barium (or strontium or calcium), with the release of oxygen. Confining our description to the case of barium, we may say that the barium atoms in the bulk of the oxide coating find their way to its surface by diffusion and form a monatomic layer which, for optimum emission, need not cover the oxide surface completely. Partial reduction of the oxide takes place at the "sinter temperature" of about 1350° K. During its use the cathode is operated at about 1000° K and diffusion of the atoms to the surface may continue during its life. This thermal activation in the absence of externally applied voltages, which we have just described, is possible only in the presence of reducing agents. It is a reliable technique for mass production and does not require any prolonged ageing of the cathode after the tube has been exhausted. Electron emission to its full expected value can be obtained from such a cathode when it is tested after completion of the activation process.

Aside from this purely thermal activation, another activation procedure may be followed which is characterized by the application of anode voltage during the high-temperature treatment of the cathode while the tube is evacuated. It may then be assumed that the oxide is separated into barium ions and oxygen ions which will migrate under the influence of the applied field according to Faraday's laws of electrolysis. The direction of the field is such that the positive barium ions drift through the oxide toward the core metal, where they give up their charge, while the negative oxygen ions migrate toward the surface of the coating and escape. The neutral barium atoms will then again diffuse through the oxide coating from the core metal to the surface and form the active surface layer, which is the origin of the electron emission under normal operating conditions. This electrolysis activation does not require any reducing agents and can thus be applied to materials of extreme purity.

The existence of the thin barium film, the way in which it is produced, and the seat of the electron emission, have been subjects of lively controversy for many years. However, very conclusive experiments have been carried out that prove the existence of the barium film at the cathode surface and show that without its presence the copious electron emission ceases; the barium film evaporates at 1600° K. The electron microscope has added greatly to a better understanding of thermionic emission, since it not only permits visual observation of the intensity distribution of the electron emission over the cathode surface during operation, but also reveals minute changes in emission from point to point, at the surface of the coating.

Comparison of the Electron Yields of Emitters

Pure metals, such as nickel, tungsten, molybdenum and platinum, do not emit electrons unless they are heated to very high temperatures. Only tungsten and molybdenum are used as "bright emitters" in practical cases. To give an example, an oxide-coated cathode will emit 100 ma. per square centimeter of its surface when at a temperature of 1000° K. It will require a

certain number of watts to reach and maintain this temperature. If we divide the total emission obtained from a cathode by the number of watts supplied to the cathode heater, we arrive at the emission yield, which is about 20 ma. per watt for an oxide-coated cathode. A tungsten filament operating at 2300° K will give about the same emission per sq. cm. as the oxide cathode at 1000° K, *i. e.*, 100 milliamperes per sq. cm., but the emission yield will be less than one milliampere per watt.

In order to produce a current between such cathodes and the anode it is necessary to apply a positive potential to the latter. Now it requires a certain potential difference at the cathode surface just to pull the electrons out of the surface. The ease with which a cathode surface is able to emit electrons is measured by this voltage, which is known technically as the *work function* for which the symbol, ϕ, is used. The lower the value of ϕ, the more copious will be the emission and the lower also will be the operating temperature necessary to obtain a certain required emission current. Table 3 gives the work function, ϕ, for a number of different cathode materials and also the practical operating temperature, the melting point, the emission yield in milliamperes per watt and the electron emissivity in milliamperes per square cm. of cathode.

Table 3.

Emitter	Symbol	Melting Point (°K)	Operating Temperature (°K)	Work Function ϕ (volts)	Emission Yield $\frac{ma}{watt}$	Electron Emissivity $\frac{ma}{cm^2}$
Barium	Ba	1123		2.52		
Strontium	Sr	1030		2.1		
Calcium	Ca	1083		3.2		
Barium Oxide	BaO	2196	1000	1.1	20.	100.
Strontium Oxide	SrO	2703	1000	1.4		10.
Calcium Oxide	CaO	2845	1000	1.9		1.
Tungsten	W	3643	2300	4.52	1.	100.
Molybdenum	Mo	2890	2200	4.15		100.
Platinum	Pt	2047		5.32		
Nickel	Ni	1725		5.03		
Thoriated Tungsten	W-Th		1900	3.15	40.	1200

Cathodes with Films of Barium

It has been found that electron emission can be obtained at comparatively low temperatures from pure metals, such as nickel, tungsten and molybdenum, when a thin film of barium is artificially deposited upon them. Since these metals have ordinarily a fairly high work function, the effect of the added barium layer must be to lower their work function.

The barium layer can be deposited in various ways. Barium vapor may be directed against the metal surface from a tiny oven in which barium metal is heated; after the required deposit is obtained, the oven is moved to the side so as not to interfere with the emitted electron stream. Instead of this little barium oven, a regular oxide-coated cathode may be used which is brought near to the metal surface to be coated. When this oxide-coated cathode is heated to about 1600° K, its surface film of barium will evaporate and deposit on the adjacent metal surface. The auxiliary cathode can then again be moved out of the way by some convenient mechanism.

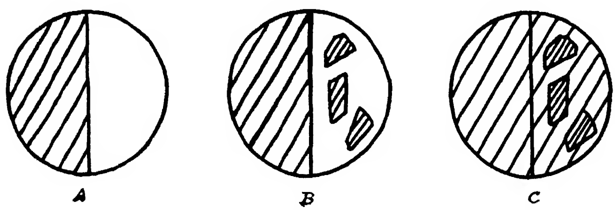


FIG. 98. Spread of areas of electron emission on a metal surface.

Finally, a metal surface, for which nickel often serves, may be only partially coated with oxide as shown by the shaded half of a circular cathode in Fig. 98a. One might then expect that electrons would be emitted only from the shaded half after the cathode has been activated. The image of such a cathode in the electron microscope will disclose however that certain areas of the uncoated half circle also begin to emit electrons at the operating temperature of the oxide cathode, *i.e.*, 1000° K. Fig. 98b is an image of such a cathode. After prolonged oper-

ation the emission will have spread to the entire half circle which was uncoated, but the first activated areas on this half will still stand out brightest (Fig. 98c). Investigations of this kind were carried out by E. F. Richter in 1933 and brought forth the following results.

The thin barium film on the surface of an oxide-coated cathode in the activated state evaporates, and beams of barium atoms are emitted in all directions as long as the supply of barium lasts. This is illustrated in Fig. 99, which gives a cross-sectional view of the cathode in Fig. 98. Most of the barium is evaporated away from the cathode and will condense on electrodes and tube walls in the vicinity. Some of the barium atomic beams, however, strike the cathode itself, *i.e.*, the uncoated section, at glancing angles and will condense on the bare nickel if this is at a

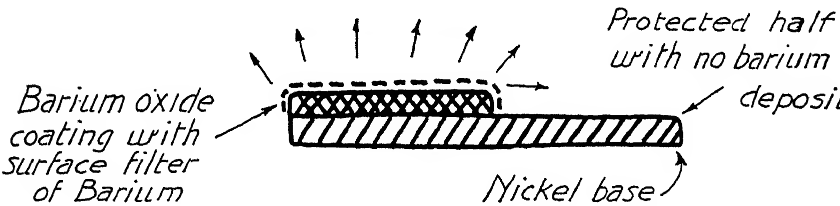


FIG. 99. The spread of atomic beams from a coated cathode surface.

lower temperature than the surface of the oxide. Such a temperature gradient exists when a reasonably heavy electron current is drawn from the oxide coating at elevated temperature. Due to the electrical resistance of the coating, the emission current passing through it will produce Joule heat that will bring the coating to a temperature higher than that of the underlying nickel by as much as 100° C. When no electron current is drawn, the nickel will have a higher temperature than the coating, since the oxide is a poor conductor of heat. In the latter case the barium deposited on the nickel will re-evaporate and the oxide surface will replenish itself with barium by diffusion from within the oxide.

Depending on the prevailing conditions, an inversion of the electron optical image of the cathode can thus be observed. The

Courtesy American Cyanamid Co.

Tungsten oxide smoke (×30,000).

oxide-coated sections of the cathode may appear bright and the bare metal dark in one case, and the opposite conditions may prevail in another, depending on the relative temperature distribution and the history of the cathode.

Rather than leaving one half of the cathode bare, as indicated in Fig. 98, the base metal of the cathode, the nickel, is readily exposed by ruling scratches in the oxide coating. When parallel scratches are ruled at measured distances on the coating the space between these scratches in the image gives a ready measure of the magnification obtained with the electron microscope. The barium will be deposited in these scratches according to the mechanism illustrated in Fig. 99 or re-evaporated from them. Fig. 100 (p. 176) gives reproductions of electron images obtained by W. Knecht in 1934 with such ruled oxide cathodes; one picture obtained by C. E. Hall at Toronto is also shown.

Disclosure of the Crystalline Structure of Metal Surfaces

Investigations with the electron microscope such as those illustrated in Figs. 98 and 99 have disclosed that barium vapor and other alkali vapors condense on a clean metal surface in such a manner as to prefer certain patches on the surface where they build up a film of condensate first, and deposit over the whole surface only after a time. These patches of irregular shape turn out to be faces of crystals of which the original metal surface is composed. Any highly polished metal surface, unless it is a single crystal, can be made to expose its crystalline structure under an ordinary microscope after the surface has been etched by suitable solutions; this technique is commonly employed in metallography. It is necessarily limited to observations at room temperature and cannot disclose changes in crystal structure that may take place at elevated temperatures or on re-cooling to lower temperatures. On the other hand the barium-activated metal surface will give not only a sharp and enlarged image of the crystalline structure of the metal without etching, which technique is apt to destroy fine detail at times, but will disclose also transitions of the crystal lattice from one type to another

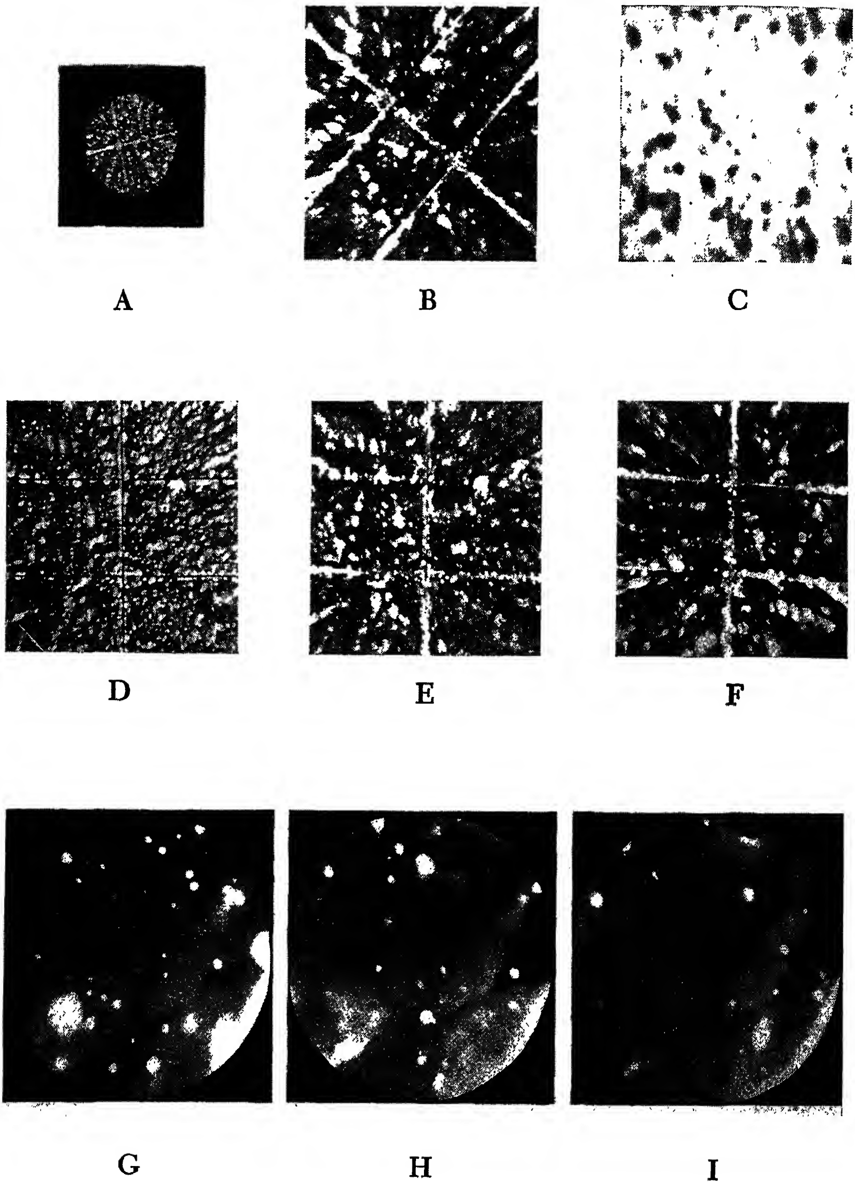


Fig. 100.

All the reproductions on this plate are of cathode surfaces emitting electrons. These cathodes are the sources of the electron beam which is focussed on the photographic plate by means of single electrostatic or magnetic electron lenses.

A, **B**, **D**, **E** and **F** are from a paper by Knecht [*Ann. d. Physik*, 20, 171, (1934)]. **A** has magnification ×9, the rest, ×31. **D**, **E**, and **F** are optical, electrostatic electron and magnetic electron microscope pictures, respectively.

C is an electrostatic electron microscope picture of an emitting cathode obtained by C. E. Hall at Toronto in 1936.

G, **H** and **I** are from a paper by Brüche and Mahl [*Z. Techn. Physik*, 16, 240, (1935)], showing electrostatic electron microscope pictures of emitting thoriated tungsten cathodes, showing the distribution of the emission points.

as the temperature is raised and lowered. The electron microscope has thus become a powerful tool in metallurgical research. Brüche and Knecht (1934) photographed electron images of the transition of iron from state to state between 615° C and 1050° C. Pictures in Fig. 101 (page 180) are reproduced from this work.

In the series of photographs in Fig. 101, the crystal structure of the iron cathode which was first completely coated with alkaline-earth oxide is laid bare after the cathode has been operated at such a high temperature that most of the oxide has evaporated and only a thin film of barium remains.

It follows from all these observations that adjacent crystals on the cathode surface emit electrons to a different degree at a given temperature, since a difference in brightness results on the fluorescent screen. That is to say that the work function, ϕ, is not uniform over the cathode surface but has specific values that differ from each other by as much as 0.3 volt for different crystal faces on the surface. Recognition of this fact has made necessary a modification of the theories on thermionic emission and has fortified a theory known as the "patch theory." Before the advent of the electron microscope the emission current measured from a cathode represented an average value, since all surface elements contributed to this current. Now, it is possible to single out a very small surface element and determine its share in the overall emission current quantitatively by means of a Faraday cage with a small aperture which is placed in front of the luminescent screen where the enlarged image of the cathode is ordinarily obtained, (Fig. 102a). D. Schenk carried out such measurements in 1935 and a plot of his observed emission currents across the cathode face, together with the electron image of the cathode strip taken separately is given in Fig. 102b and c. The trace along which point to point emission readings were taken with the Faraday cage is indicated by the dark line in Fig. 102c. Thus it is clearly evident that large emission currents result from crystal faces that appear bright in the electron image.

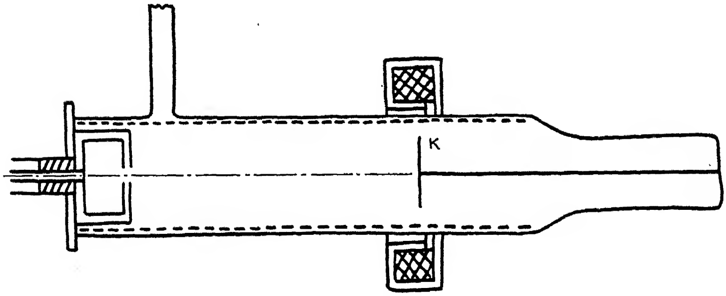


FIG. 102a. The number of electrons emitted by a small area of the cathode, K, is measured by having the aperture of a Faraday cage catch the electrons which would fall on the corresponding small area on the image if not intercepted.

FIG. 102b.

Graph showing charge given by Faraday cage as its aperture was passed along the line shown in Fig. 102c.

FIG. 102c. The electron image of an area of a cathode showing patches corresponding to varying emission of electrons.

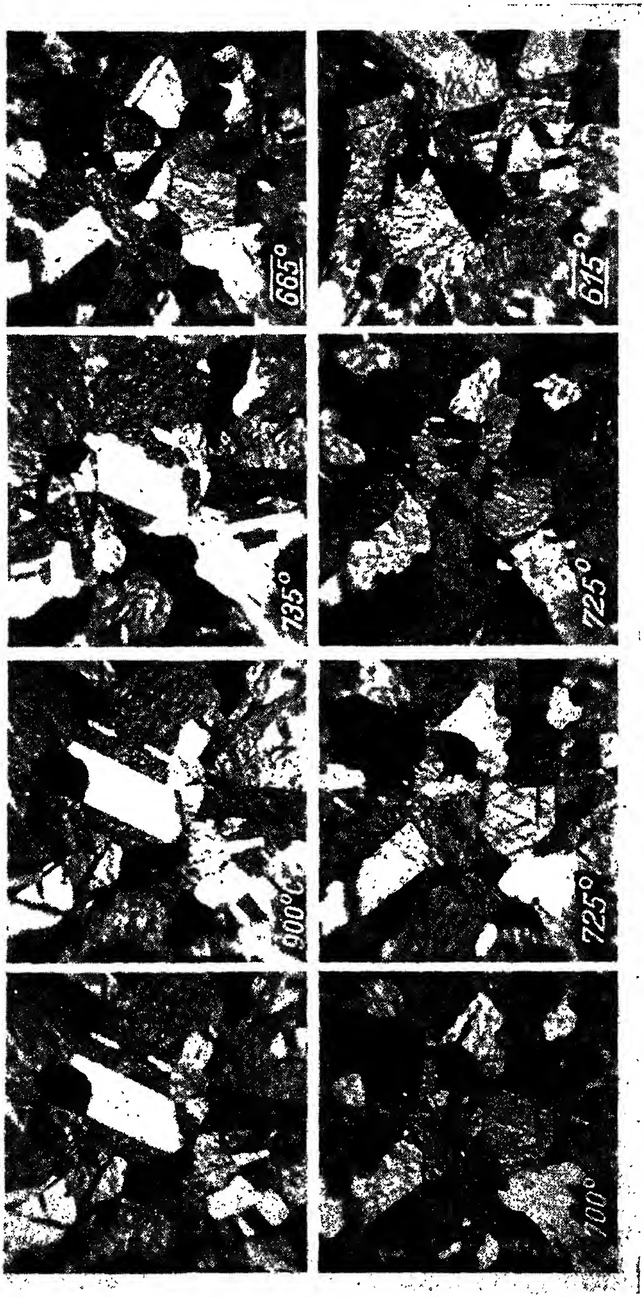


FIG. 10

FIG. 101. Reproduction from paper by Brüche and Knecht (1934) giving the electrostatic electron microscope picture of the surface of an iron cathode which was covered with barium oxide and then operated at such a high temperature that most of the oxide coating has evaporated, leaving only a thin film of barium. The cathode was then taken through a series of temperature changes and pictures taken at various temperatures. The magnification was not stated in the paper [Z. *Techn. Physik*, 15, 462, (1934)] but is probably a few hundred times.

The bright areas are regions of high emission. The intensity of the emission is thought to be determined by the distribution of barium on the surface. Small crystal sections are shown on the surface by the brightness of the emission.

The iron was heated to 1050° C. for the first picture (lower left) and then the temperature was allowed to fall to 900° C and again raised to 1050° C, but there was no apparent change in the crystal arrangement. The temperature was again dropped during a period of five minutes to the low temperature of 735° C. On again raising the temperature, no remarkable changes were apparent. A succeeding lowering of the temperature to 665° C led, however, to a complete reorientation of the structure. Raising the temperature to 725° C again and again did not change the arrangement essentially. Lowering the temperature to 615° C, however, once more changed the picture completely. This action of iron is paralleled by similar changes in nickel.

Emission Patterns on Wires

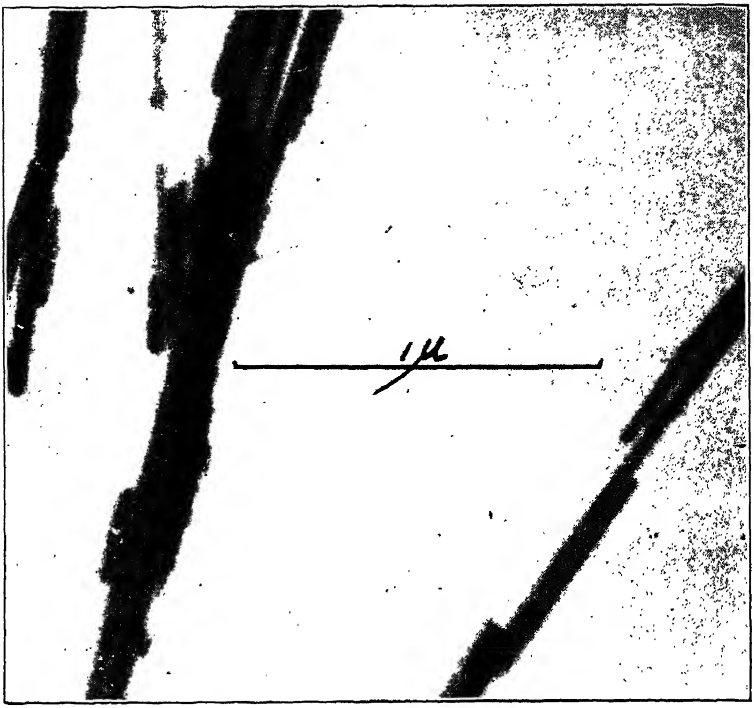
Similar investigations on emission patterns have been carried out for bright emitters which are generally used in the form of wires. The effect of die marks on drawn filaments becomes plainly visible in electron images, as R. P. Johnson and W. Shockley demonstrated in 1935. With a tungsten filament 0.005″ in diameter stretched along the axis of a cylindrical tube, the inside wall of which was coated with luminescent material and onto which a helix of nickel wire fitted snugly as positive electrode, a lateral magnification of the wire surface equal to 400 times was obtained with 10,000 volts on the anode spiral. This gave a resolution of points on the filament surface 2° apart. Along the axis of the wire naturally no magnification is obtained with this simple arrangement.

A problem of particular interest, for a long time, has been the determination of the mode of activation of thoriated tungsten filaments. When a small amount of thorium oxide (ThO_2) equal to about two per cent is mixed with the tungsten powder during the manufacture of the wire, a substantial reduction of the work function, ϕ, can be obtained after proper activation of the filament. This activation is similar in principle to that of the oxide-coated cathodes. The filament is "flashed" at 2700° K for about 30 seconds and the temperature then lowered to 2100° K and held there for some time. The filament is then operated at 1900° K during its life. At 2700° K part of the thorium oxide (ThO_2) is reduced to metallic thorium and oxygen (which escapes). The thorium diffuses through the crystal lattice and evaporates from the surface of the wire. When the temperature is lowered to 2100° K, the thorium which arrives at the surface will spread over it, forming a monatomic layer. The effect of this thorium layer on tungsten is similar to that of barium on the barium oxide layer on nickel. It lowers the work function and thus increases the electron yield at a lower operating temperature. Table 3 contains the characteristic figures for thoriated tungsten filaments.

With the aid of the electron microscope, the arrival of the thorium on the tungsten surface can be observed and the spreading of the thorium patches over the surface followed. An earlier assumption that the thorium atoms arrive at the surface preferentially along crystal faces was thus disproved by Brüche, Johannson and Mahl, and their conclusions were confirmed by A. J. Ahearn and J. A. Becker in 1936. Pictures G, H, and I on page 176 are reproductions from a paper by Brüche and Mahl showing the appearance of thorium centers indiscriminately over the surface, some on crystal boundaries, some right within a crystal face.

The number of applications that suggest themselves and which have been carried out in the past is very large. We shall here go no further in their description since this book is not to be a technical text but rather an introduction to the subject of electron microscopy for the reader who has heretofore not been acquainted with it. The specialist will be familiar with technical publications in the various journals.

It should be said, in closing this chapter, that various problems of research that may be attacked with the electron microscope cannot always be definitely assigned to either the electrostatic or the magnetic electron microscope and that the investigator will frequently have a choice of either type of instrument. Some of the examples quoted above could thus equally well be treated under the heading: Applications of the Magnetic Microscope. The investigation of filament surfaces has mostly been carried out with the magnetic microscope. The latter instrument is a much more costly piece of apparatus but more flexible in its use. We shall now turn to its description.



Abestos fibers (×46,600).

Edge of pollen dust particle (×21,000).

Chapter 16

The Compound Electron Microscope—Magnetic Type

The Comparison of Light and Electron Compound Microscopes

In preceding chapters we have indicated how an electron beam has been used to simulate effects produced by ordinary light in the simple microscope or reading lens. Instead of a glass lens as in the light microscope (see Fig. 15, p. 26) we can use either an electrostatic field (Fig. 58) or an electromagnetic field (Fig. 82). In the case of the simple electron microscope the object is always the source emitting the electrons, viz., the cathode of a vacuum tube. As a consequence the chief applications of this form of electron microscope are the studies of surfaces on which electron emitting materials have been deposited. These applications have been quite fully illustrated.

We now come to deal with the development in electron microscopy in which the electron beam from such a source is used to reveal the intimate structure of quite an independent object. Since the whole purpose of such a development is to produce magnifications greater than the ordinary light microscope can possibly give, the problem before us is to simulate the performance of the ordinary compound microscope, shown schematically earlier in Fig. 16 (p. 27) and again in Fig. 104, a and b.

In its ordinary use, the compound microscope is used to examine thin sections which are partially transparent; the thicker part of the specimen is illuminated by the light source and the light diffused from the object is focussed on to the retina of the observer's eye or the photographic plate. The aim of good light microscopy is to so illuminate the object viewed as to make it as

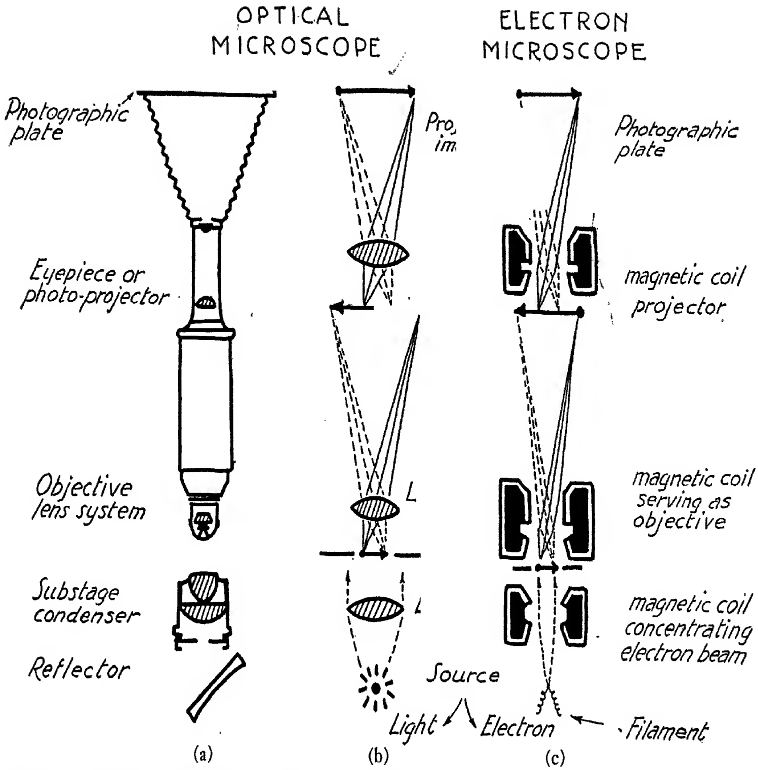


FIG. 104. Comparison of light microscope and magnetic electron microscope. (See Table 4, opposite.)

Table 4. Comparison of Light and Electron Compound Microscopes. (See Fig. 104, opposite)

Elements	Light microscope	Electron microscope
1. Source of illumination.	Sunlight or electric light.	Electron source: cold cathode or hot filament.
2. Control of illumination.	Sub-stage condenser, L_1. Fig. 104b.	Converging action of the magnetic field of the first coils.
3. Specimen support.	Microscope slide of glass about 1 mm. thick.	Film of collodion about 10 millimicrons thick.
4. First image forming system.	The objective, L_2.	Converging action of the magnetic field of the second coil forming the first image.
5. Accessory magnification.	The eyepiece or photographic projector, L_3.	Converging action of the magnetic field of the third coil forming a second enlarged image of a portion of the first image.
6. Medium.	Air and glass	A very high vacuum.
7. Viewing image.	The eye directly.	The eye (image projected on fluorescent screen.)
8. Method of focussing.	Movement of lenses as a result of judging the sharpness of the image with the eye.	Alteration of the intensity of the magnetic fields in the coils as a result of judging the sharpness of the image on a fluorescent screen.
9. Recording of image.	Photographic plate.	Photographic plate.
10. Smallest particle photographed in detail.	100,000 cm.	Even less than $\frac{1}{1,000,000}$ cm.

nearly as possible self-luminous. As indicated in Figs. 104a and 104b, the essential features of the ordinary microscope are: the light source, the sub-stage condensing lens to concentrate the light on the object, the object itself on the stage, the first set of lenses known as the objective, a second set of lenses some distance from the objective known as the eye-piece (for observation with the eye) or as the projector lens when it is designed to focus the image on a screen or a photographic plate.

As indicated in Fig. 104c, all of these elements are present in the electron microscope. Table 4 (page 187) shows the correspondence between the two instruments.

It will be apparent from what has been presented in the preceding chapters that if we should write out a description of an ordinary compound light microscope and describe how it works, we would obtain almost a complete description of the present electron microscope merely by replacing the word "light" by "electron beam," the word "lens" (glass) by the words "magnetic lens."

Precise Designation of Useful Magnifying Power

The relationship between resolving power and magnifying power for an ordinary optical system has already been dealt with rather fully (Chap. 2) and we may now state that for both the electron microscope and the light microscope the detail which is visible in the final image is determined by the objective lens. In each case the objective produces an image which contains all the detail that will be visible in the final picture. However, the image produced by the objective is seldom at a magnification of more than $\times 100$ in either instrument and hence most of the detail present still cannot be seen by the eye. The subsequent magnification by the eye-piece in the case of the light microscope, and the projector coil in the case of the electron microscope, merely serves to increase the dimensions of these details until they are visible to the eye. In order to be visible, the finest detail of the picture must have a linear dimension larger than 0.01 cm. or 1/250th of an inch.

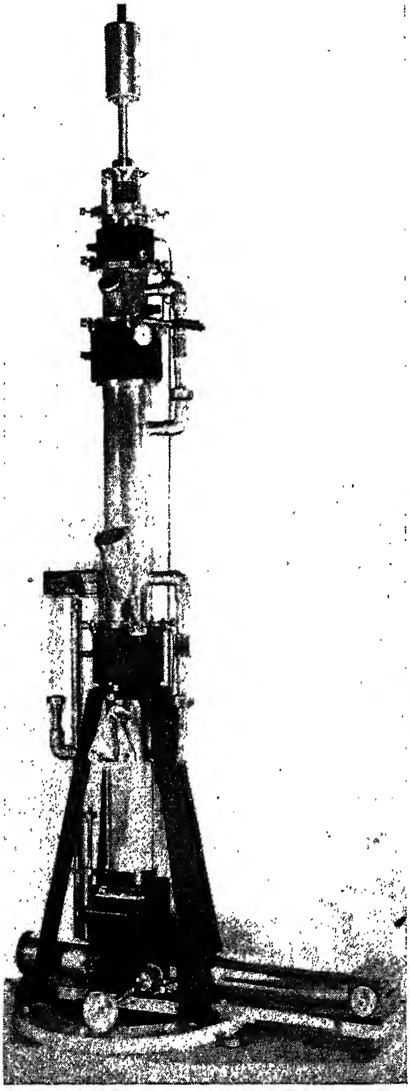
In Chapter 2 we have pointed out that due to the limitations of the wave length of light no light microscope can possibly reveal to us a particle or detail of dimensions less than about 1/50,000th cm. (or 1/125,000th of an inch). If then a detail of this dimension is in the image made by the objective, the magnification necessary to be provided by the eye-piece of the microscope is $\times 5$; that is, the over-all magnification is $\times 500$ and 500 times 1/125,000th of an inch gives a dimension 1/250th of an inch, which will just make this detail visible in the image. In practice a magnification of $\times 2000$ tells everything; one can go on enlarging and enlarging the photograph but get no more detail (see p. 30).

Now with the electron beam, the wave length is so much smaller that we are not limited to dimensions of 1/125,000th of an inch as in the case of the light microscope. We find consequently that in the image obtained by the objective of the electron microscope alone there is detail which requires much more magnification by the projection lens (equivalent to the eye-piece) than is the case with a light eye-piece or light projection lens. Instead of producing a mere $\times 20$ magnification, the objective image will stand $\times 200$, or even more. That is, a very ordinary final picture will have a magnification $\times 20{,}000$. As will be apparent from many of the plates, these final electron microscope pictures will require and permit enlarging up to $\times 100{,}000$ in order to display all the detail intrinsically in the original picture.

Description and Manipulation of the Instrument

Fig. 105a is a photograph of the first Toronto instrument and Fig. 105b gives the skeleton structure. The description of the instrument as given in an early paper is as follows:

"Electrons leaving the cathode, *F*, are accelerated vertically downwards and pass through the anode, *A*, in the form of a narrow beam of uniform velocity. The magnetic field of the condenser lens, *Co*, converges the beam to a small cross-section in the plane of the object at *O*. The elements of the object scatter the electrons incident on them. The scattered electrons, which



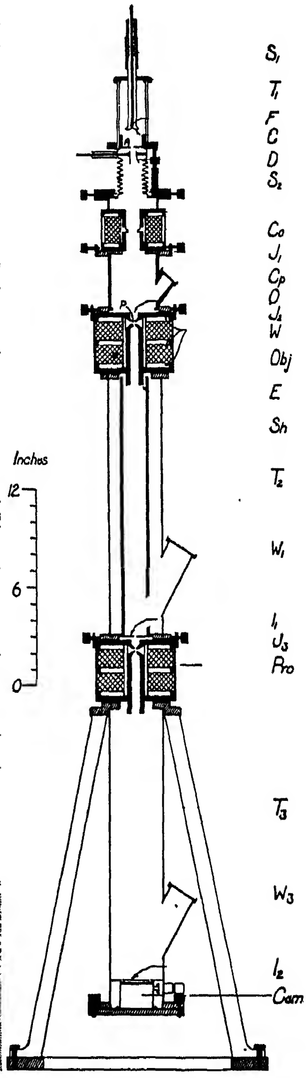


FIG. 105a FIG. 105b

Photograph and diagram of the electron microscope now in use at the University of Toronto. The description is given in the text.

on leaving the object pass through the objective lens, *Obj*, are focussed by the magnetic field of this lens to conjugate image points upon a fluorescent screen, I_1. The electrons from any desired part of the image are directed through an opening in the screen into the projector lens, *Pro*. They are focussed by this lens to form a magnified image of the selected part of the first image on the final fluorescent screen, I_2. The image may be recorded photographically by replacing the screen with a photographic plate. A high magnification may be obtained in each stage by choice of a large ratio of image distance to object distance. The magnification in the first stage lies between 100 and 125 diameters and is determined by the vertical displacement of the specimen relative to the objective lens. The magnification in the second stage may be varied over wide limits (from $\times 1$ to $\times 330$ at 50 kV) by varying the focal length of the projector lens and altering by small amounts the vertical displacement of the first image relative to the projector lens. The total magnification is the product of the magnifications in each lens, provided the first image plane lies above the region in which the magnetic field of the projector lens is appreciable.

"The apparatus is constructed in several units which are sealed together by means of plane lapped grease joints. The discharge tube, *T*, the bellows adjustment, S_2, and the condenser lens, *Co*, are included in the top unit. It may be adjusted in any direction in the horizontal plane by means of the joint, J_1, and fixed in position by means of four set screws. This adjustment together with the bellows, S_2, permits the axis of the illuminating electron beam to be adjusted both in vertical inclination and laterally with respect to the axis of the object chamber and objective lens below.

"The object chamber is the second unit. It is sealed to the top of the objective lens, *Obj*, by means of the grease joint, J_2. It may be adjusted laterally with respect to the objective lens and fixed in position by means of a third set of four set screws. Specimens are inserted into the apparatus through the opening, *Cp*, of

the object chamber. This opening is a conical grease joint. The specimen holder is in the form of a hollow cartridge which fits snugly into a collar attached rigidly to the case of the object chamber. The lower end of the cartridge is about 5 mm. above the centre of the gap between the pole pieces of the objective lens. This arrangement makes it possible to replace the specimen holder, to within less than a tenth of a millimetre laterally and a few one hundredths vertically, in the position it occupied before removal and reloading. The specimens are suspended across a small hole (0.05 to 0.3 mm.) drilled in the centre of a circular metal diaphragm, or across the openings of a diaphragm made out of fine mesh. The diaphragm is held at the lower end of the specimen holder by a cap. A displacement of the object chamber with respect to the lens displaces the specimen by an equal amount.

"The objective lens, *Obj*, forms the third unit. Its lower side is sealed with grease to the top of the first image tube, T_2.

"Concentric with and extending along the entire length of the brass tube, T_2, are two soft iron cylinders, *Sh*. These are approximately one, and one and one-half, inches in diameter respectively, with a wall thickness of one eighth of an inch. They shield the electron beam in the tube from the perturbing effects of stray fluctuating magnetic fields. The tube, T_2, is the fourth unit of the structure. It is sealed to the top of the projector lens with grease and may be adjusted laterally with respect to it by means of a flat grease joint, J_3, and set of four set screws.

"The fluorescent screen, I_1, which is shown in Fig. 105b as fixed in the base of the tube, has been mounted on the top of the projector lens, *Pro*. This permits the first image to be displaced relative to the projector lens by a small amount (up to 3/8") and the displacement may be measured by observation of the motion of the image relative to marks drawn on the screen. The screen may be viewed through the window, W_1.

"The projector lens is the fifth unit of the apparatus and it is sealed with grease to the top of the tube, T_3.

"The base of the apparatus consists of a heavy iron ring, upon which the feet of four bronze legs are bolted. The upper ends of these legs are bolted to a bronze ring, the upper side of which is lapped and which supports the projector lens and the sections above it. The upper end of the tube, T_3, is soldered into the ring and its lower end into the camera.

"The apparatus is evacuated through a connection in the side of the tube, T_3, by means of a mercury diffusion pump backed by a Cenco Hyvac fore-pump. By-passes between the object chamber and first image tube, and between the first and second image tubes (see Fig. 105a) facilitate rapid evacuation of the entire apparatus.

"The source of the current for the lenses consists of a set of four standard 12-volt storage batteries connected in series. The cathode filament current is supplied by two 2-volt storage batteries in series; these are mounted in an insulated box which forms part of the high tension system.

The Discharge Tube

"The schematic cross-section shown in Fig. 106 illustrates the construction of the cathode of the discharge tube. The electrons are emitted from the filament, F, which consists of 13 mm. of 6 mil tungsten wire bent in the form of a hairpin with a radius of curvature of 0.5 mm. at the tip, T. The ends of the filament are welded to heavy nickel wire supports, L_1, L_2, which are fastened with screws into two collars, B_2, B_3, mounted at the end of a long vertical rod, R. The support, L_1, makes contact with the rod, R, through the collar, B_3. At the upper end of the rod is fastened one of the exterior terminals, T_1. The nickel support, L_2, is fastened in the collar, B_2, which slides in a vertical direction within the tube, D, and which is insulated from the central rod, R. A spring connection of heavy copper wire, S, ensures electrical contact between the collar, B_2, and the body of the cathode. Contact with the battery is made through the exterior terminal, T_2. The rod, R, may be displaced vertically with respect to the body of the cathode with the aid of the bellows, A, and the

adjusting screws, *N*. The wax seal and mica bushing at *W* together with the glass tubing insulate the rod from the main body of the cathode. A water jacket, *C*, is provided to cool the top of the cathode. The cooling is necessary to prevent the wax seals from melting and also aids in the rapid establishment of temperature equilibrium after the filament heating current is switched on.

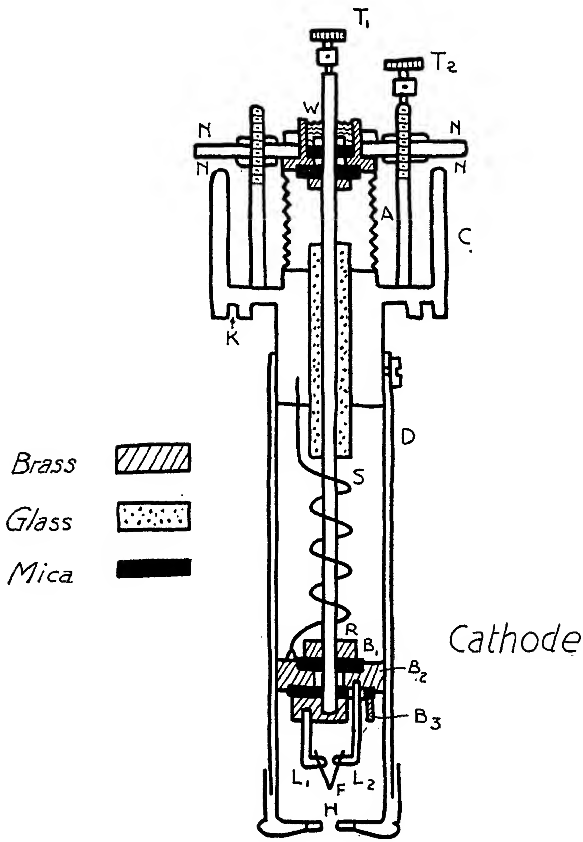


FIG. 106.

Detailed drawing of the cathode of the discharge tube in the magnetic electron microscope.

"The anode consists in part of a copper cylinder the top of which extends around and above the lowest extremity of the cathode. The extension is required to shield the electron beam from the effects of charges built up on the glass insulator. The

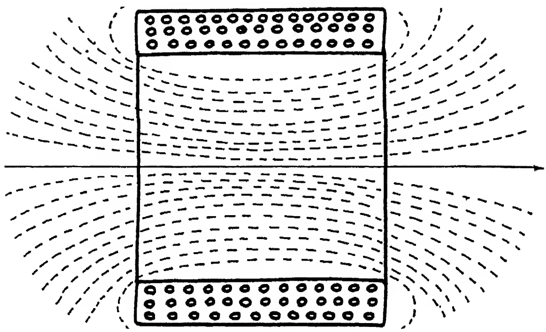
base of the anode cylinder consists of a copper disc in which holes are drilled. A central hole is provided as a channel for the electron beam, and a number of large holes with centres on a circle concentric with the centre of the disc, are provided to allow rapid evacuation of the discharge tube. The entire discharge tube is mounted on the upper end of the bellows, S_2, (Fig. 105b) which permits the inclination and lateral position of the tube to be adjusted with respect to the condenser lens." (Dr. A. Prebus)

The Iron-Shielded Magnetic Lens

The lenses of this form of electron microscope are all magnetic electron lenses of the type described in Chapter 12; that is, short solenoids mounted coaxially with the axis of the instrument. They are, however, not open solenoids as indicated in these earlier diagrams but special iron clad coils with only an annular zone where the inner enclosure is of non-magnetic material. Since these coils are a principal part of the magnetic electron microscope, we shall give a little more detailed description of them.

When speaking of lenses in ordinary optical instruments a distinction is made between "thin lenses" and "thick lenses." These terms are used not only in their everyday meaning of thin and thick but also to indicate that a different method of ray tracing or calculation must be followed in the two cases in order to determine the properties of the single or compound lens respectively. A lens is called thin when its thickness along the axis is a small fraction of its diameter. This means in turn that the light ray travels only for a short fraction of its entire path through the medium of higher refractive index.

In the light of this terminology the introduction of Busch's short solenoid instead of the long solenoid was a step towards a thinner lens. But in optical language this Busch lens is still a thick lens since the magnetic field strength along its axis is a considerable fraction of the maximum intensity at the center of the coil for quite a distance from the center of the coil on each side. Fig. 107 illustrates this condition.



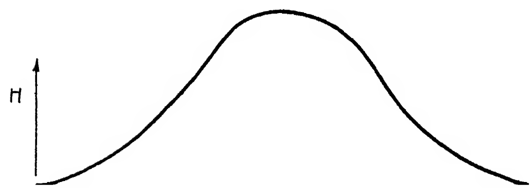


FIG. 107. Busch's original short solenoid magnetic lens. The graph shows the distribution of field strength along the axis.

It was recognized by D. Gabor in 1927 that the magnetic field could be concentrated along a shorter distance along the axis of the coil if the latter was encased in iron as shown in Fig. 108; the cross-shading here indicates the winding and the solid black frame the soft iron enclosure. It is evident that the stray magnetic field is reduced considerably since the magnetic field lines prefer to run through the iron casing.

M. Knoll and E. Ruska in 1931 carried this restricting influence of the iron shield to the extreme and utilized the stray field at a narrow gap as the active field of the magnetic lens. Fig. 109a shows this type of coil in its earlier form and the field pattern obtained with iron filings. Fig. 109b gives the field distribution along the axis of the coil.

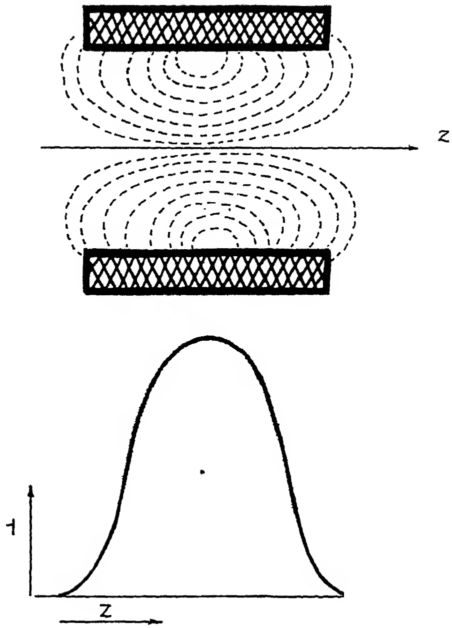


FIG. 108.

Solenoid encased in soft iron to concentrate the field at the center.

It was found that a reduction of the coil current by 30 p.c. results from this type of construction, that is, a required field strength can be obtained with about one third of the coil current that would be necessary in an unshielded coil. The distribution shown in Fig. 109b was obtained when a current of 100 ma. was sent through the coil shown in Fig. 109a. The winding of this coil consisted of 6500 turns of 0.2 mm. diameter enamelled wire and it was operated from a 220-volt source. The focal length of this coil was of the order of a few centimeters. As a general rule it is stated that the minimum focal length of such a coil as shown in this figure is from one third to one quarter of the inside diameter of the iron shield.

The focal length of a magnetic lens is controlled by the intensity of the current which flows through the winding; it can thus be varied within a considerable range. The minimum value of the focal length, which should be as small as possible if large magnifications are to be obtained, is determined by the geometry of the coil. Thus, when the physical dimensions of a magnetic

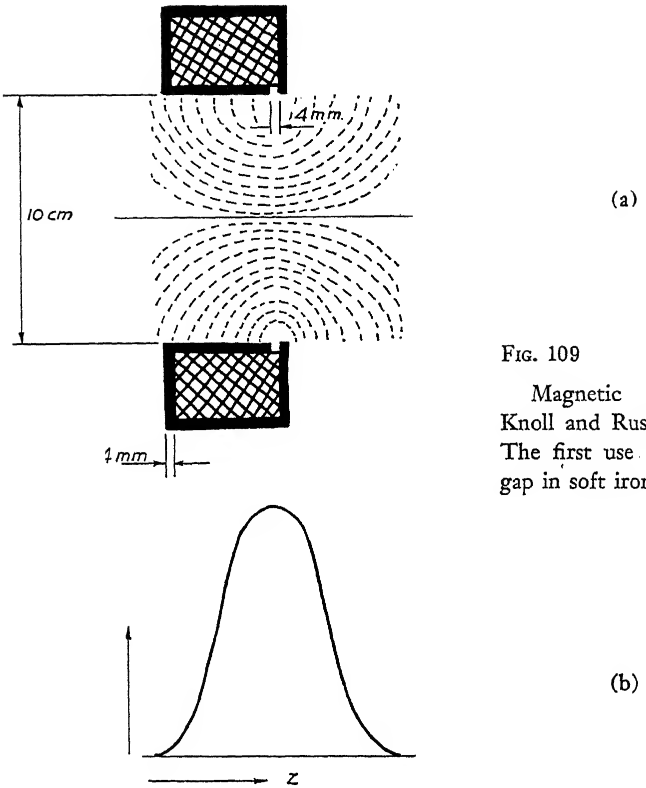


FIG. 109

Magnetic lens of Knoll and Ruska (1931). The first use of narrow gap in soft iron casing.

lens are reduced to one half of their original size and the coil current is kept constant, then the power of the lens will be doubled or the focal length reduced by half. As the magnetic field intensity, H, depends on the number of ampere turns ($n \times I$, number of turns times current in amperes) it is necessary to keep ($n \times I$) constant for the smaller coil. Thus reduction of the coil to one-half its size means that the current I will flow through a wire which has only one quarter the crossectional area, since in order to allocate the same number of turns in half the space a reduction of the wire diameter by one-half is also necessary; this will lead to overheating of the wire if the reduction is carried too far.

E. Ruska has overcome this difficulty (1934) by the introduction of pole pieces which are inserted in the threaded inner wall

of the coil and in turn contain an annular gap made of non-magnetic material such as brass where the stray magnetic field will protrude and provide a concentrated magnetic field near the axis of the coil. The physical size of the coil itself can thus be kept conveniently large so as to prevent overheating of the winding but the field is concentrated into a very small zone by the use of accurately machined pole pieces of very small dimensions. A magnetic lens of this type is shown in Fig. 110 in a schematic diagram which does not contain the details to be found in a modern

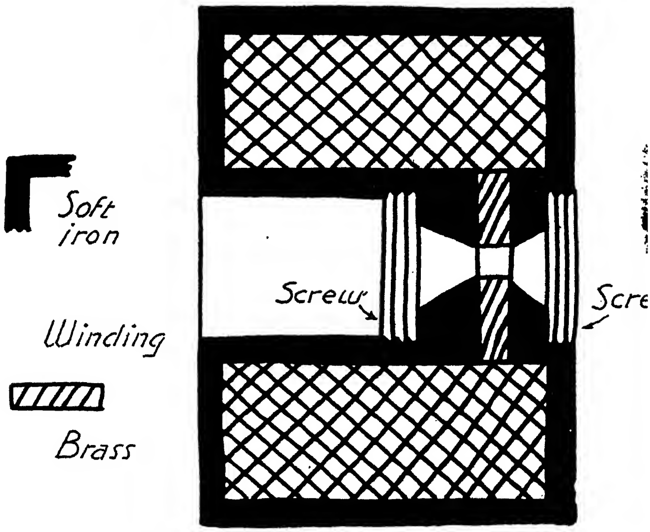


FIG. 110. Modern magnetic lens as used in magnetic electron microscope.

magnetic lens. The correct shaping of the pole pieces is a matter of great importance and has a decided effect on the quality of the image.

The three magnetic lenses used in the electron microscope, that is, the condenser lens, the objective lens and the projector lens, are all essentially of the construction shown in Fig. 110. There are, however, differences in detail which are dictated by their different functions. Thus the condenser lens is generally of the type shown in Fig. 109a while the pole piece construction of Fig. 110 is used on objective coil and projector coil where large

magnification is required. The latter two lenses differ only by the size of the internal diameter of the pole pieces and the diameter of the limiting aperture.

During the early stage of development most magnetic lenses, for microscopes operating at electron velocities equivalent to 50,000 volts or more, had to be watercooled in order to prevent overheating of the coil winding. Since the input to the coil has been reduced to about 5 watts, watercooling is not now required.

The direct current supply for the activation of the coils must be designed so as to reduce current fluctuation to a minimum. Such fluctuations, if they do exist, have the effect of shifting and rotating the plane in which a sharp image is formed. Since this final image plane is made to coincide with the photographic plate or a fluorescent screen at the bottom of the instrument when the original lens currents are chosen, a drift or a periodic fluctuation of these coil currents will blur the image. Variations in velocity of the electrons which are incident upon the object would likewise shift the focal length of the magnetic lens even if the coil current were absolutely constant because the focal length of the coil varies in a manner which is directly proportional to the accelerating potential of the electrons.

Some Experimental Difficulties

The duty of the physicist at the present stage of the development of the electron microscope is simple. He has to take hundreds of pictures of all kinds of specimens submitted by specialists in other fields: the commercial industrialists as, for example, the manufacturers of rubber constituents, paints, paper, or the medical scientist or biologist who is interested in such submicroscopic structure as the various fine organisms reveal. It is not for the physicist to interpret his results; his work is accomplished when his instrumental technique is so developed that he can vouch for the reality and dependability of his pictures.

To accomplish this work, however, the physicist has to overcome many difficulties, some of which may be enumerated here.

(1) Since, in order to have the electron beam travel any appreciable distance, one must maintain a very good vacuum, the experimenter has on his hands a continual battle against insidious leaks. This is no mean accomplishment when one considers the numerous joints and contacts between carefully ground plane surfaces, all of which have to be kept air-tight.

(2) On account of the fact that the filament is extremely sensitive to small traces of impurities, it is of utmost importance to keep the microscope quite free from vapors of such things as oils, greases, waxes, or pastes. Once such contamination is introduced, it takes some time to get rid of this difficulty.

(3) The necessity of lining up very exactly the electron beam with the axis of the instrument, which of course must coincide with the axis of the magnetic fields of the lenses, demands a very extensive series of adjustable joints, which in turn makes the maintenance of the vacuum more difficult.

(4) It has already been pointed out that the property which corresponds to the index of refraction of the glass of the lenses in any ordinary microscope, which in turn depends on the wave length of the light, is the velocity of the electrons. To insure any approach to proper focussing of the electron microscope, the velocity of the electrons in the beam must be uniform and constant. This requires a constant D. C. high voltage at the source where the electrons are accelerated. In the instrument here described this voltage is maintained at about 45,000 volts (45 kilovolts) kept steady to *one volt* during the exposure.

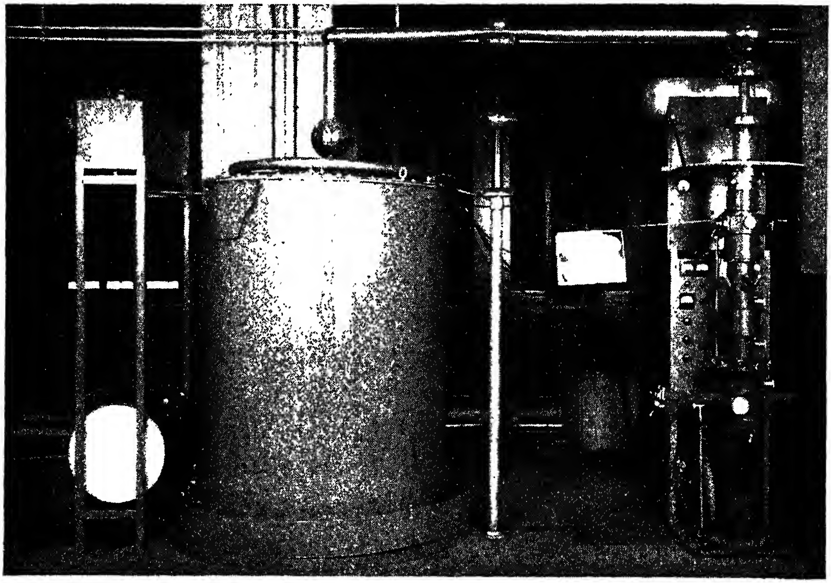
(5) Special precautions must be taken to insure that the currents through the various electromagnets are kept constant during exposures.

(6) Under ordinary routine the filament has to be replaced fairly often and easy exchange of filaments must be provided for.

(7) Continued operation of the microscope brings about a fouling of some of the critical surfaces, such as the apertures of diaphragms. Even extremely small aggregates of foreign substances may absorb electrons, create a local charged mass which

will superimpose an irregular electrostatic lens on the regular magnetic lens, and so create a disastrous distortion of the image.

It is quite beyond the scope of this book to treat at any length of the means by which all such difficulties are overcome. Such details can be obtained from more technical textbooks or from the original published papers.



New 300-kilovolt electron microscope at RCA Laboratories, Camden, N. J.

Future Developments

An Electrostatic Compound Microscope. Bearing in mind that electrostatic lenses as described in previous chapters can be made to accomplish the same thing as magnetic lenses, it is to be expected that a compound electron microscope, in which some, or even all, of the magnetic lenses may be replaced by electrostatic lenses, may be expected to be introduced before long. Advantages may be cited for each of the two forms, but there is no doubt that at the present time the magnetic compound electron microscope holds the field.

Whatever the commercial market may develop, there is not the slightest doubt that a compound electron microscope of some form will be an essential feature of every up-to-date scientific laboratory which involves microscopy to any degree.

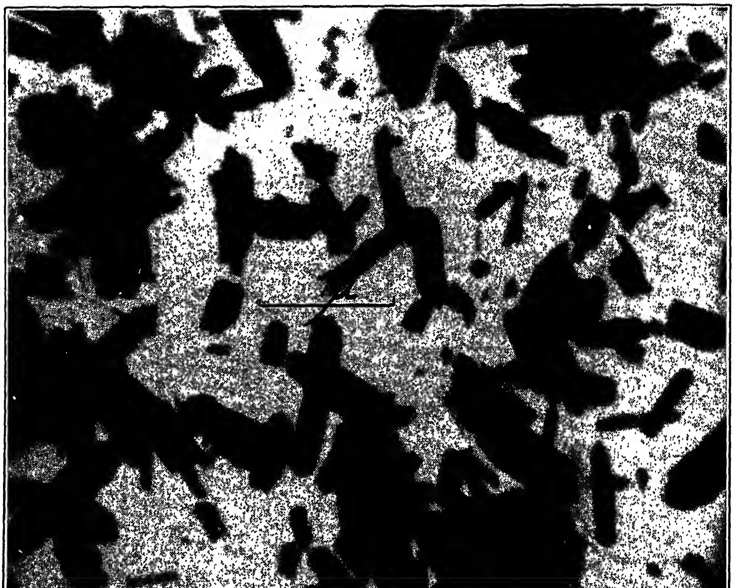
Use of Higher Velocity Electron Beams. In the instrument described here, the accelerating voltage is 45,000 volts. The higher the voltage, the higher the velocity of the electrons, the higher the "index of refraction" and the greater the penetrating power of the electron beam.

"Unfortunately, in the images obtained with the present types of instrument using electrons accelerated by from 30 to 100 kilovolts, those parts of the specimen which are thicker than about 5000 Ångstrom units (5×10^{-5} cm.) in the case of organic materials, and thicker than a few hundred Ångstrom units in the case of heavy metals, appear nearly opaque and completely lacking in detail. As many important problems of biology involve the investigation of fine structure which is an integral part of a relatively large body, and which cannot be removed from that body without loss of identity, this lack of penetration is probably the most important limiting factor in the present types of electron microscopes.

"In order to investigate the possibility of obtaining images of the internal structure of thicker specimens by the use of higher voltages for the acceleration of the irradiating electrons, the authors have constructed in the RCA Laboratories an experimental magnetic electron microscope capable of producing images of high resolving power with 300-kilovolt electrons. Except for the design of the high voltage generator and a few modifications of the microscope column resulting from the change to higher accelerating potentials, the instrument is the same as that which has already been described. . . .

"The preliminary results indicate that the expected increase in penetration is the only advantage obtained from the use of higher accelerating potentials and that the increase in penetration becomes useful only at potentials above 150,000 volts (150 kilo-

volts). There appears to be no advantage in using the high-voltage microscope to examine specimens which are transparent for the lower voltage instrument" (Zworykin, Hillier and Vance).



Iron oxide particles (×18,400).

Chapter 17

What the Electron Microscope Can Accomplish

With the plates that have been presented in the earlier chapters to illustrate certain specific points and those which appear in this concluding chapter, we have very convincing evidence of the fact that this new instrument has opened up a vast new field to man. Where not otherwise stated the pictures have been taken at the University of Toronto; some of these are among the very first produced there, and thus are among the first produced in North America.

We wish also to acknowledge gratefully the permission obtained from other laboratories to reproduce some of their photographs which show excellent technique. Through the kindness of Dr. C. E. K. Mees and Dr. A. L. Schoen we have permission to publish the remarkable pictures of photographic emulsions taken by Mr. C. E. Hall in the Research Laboratory of the Eastman Kodak Company. The R. C. A. Laboratories have kindly consented to the reproduction of various excellent photographs taken by Dr. James Hillier and his associates. Some striking photographs have been generously contributed by the American Cyanamid Company.

In this connection we wish to refer again to the support of the Columbian Carbon Company of New York City; their bro chure on the size of carbon particles is probably the first extensive publication in English dealing with electron microscope photographs.

The remaining photographs are presented under the following three headings:

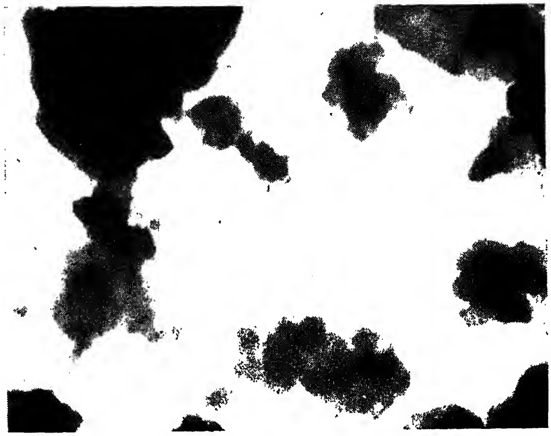
a. Measurement of particles in dusts, powders or smokes.
b. The pictures of photographic emulsions.
c. Photographs of materials with large complex molecules or molecular aggregates.
d. Biological and medical specimens.

Photographs of Dust, Smokes and Powders

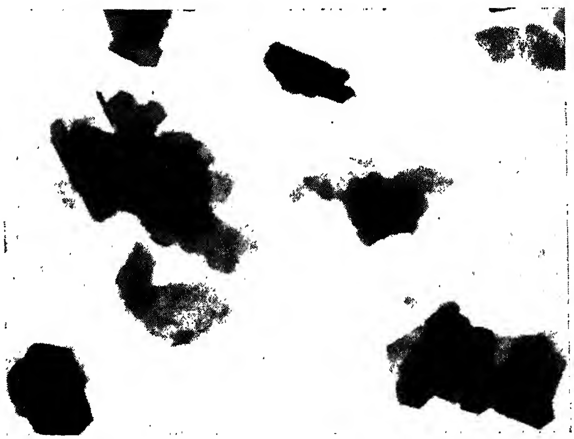
This work has extensive application both commercially and medically. Manufacturers of carbon black, fillers of various kinds, paints, and clays are discovering wonderful and surprising facts regarding the ultimate structure of their products. This work bids fair to offer an enlightening explanation of many puzzling activities of these various products.

This division of the work also has its medical bearing in the study of air conditioning, industrial diseases due to dust in factories, and the scourge of silicosis in mines.

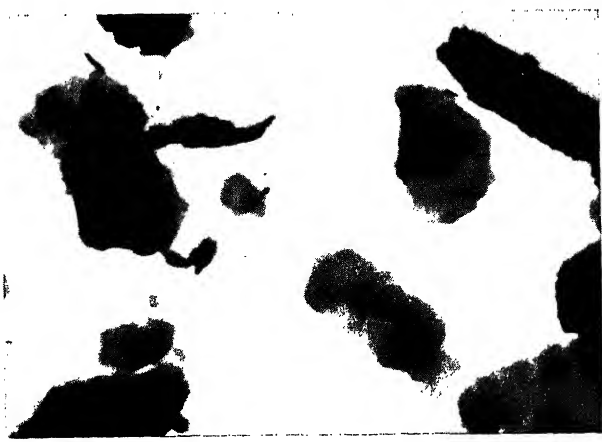
Mine dust from blasting (×14,000).



Washed St. Remi clay (Quebec) (×11,300).



Cornwall paper-coating clay (×12,000).



Georgia paper-coating clay (×12,500).

In the study of the particle size distribution in samples of smokes perhaps seventy-five per cent of particles are so small that they are completely invisible in the highest-power light microscope. Reliable distribution statistics, then, demand the use of the electron microscope.

Plates have been shown of the distribution of particles in commercial production: carbon black (p. 151), smokes of magnesium oxide (p. 144), tungsten oxide (p. 174), zinc oxide (p. 112), iron oxide (p. 204), and aluminum oxide (p. 150). Pictures of particles in various types of clays and one of the dust actually produced in a mine by blasting are shown herewith. Entirely apart from their practical or scientific interest some of these pictures are engrossing as mere examples of nature's display of shape and form.

The Eastman Kodak Pictures of Photographic Emulsions

Most of the figures reproduced here (p. 210) were published in an article by C. E. Hall and A. L. Schoen in the *Journal of the Optical Society of America* (**31**, 281, 1941) and some were sent to us direct from the Eastman Research Laboratory.

One of these plates, A (×21,500), is the picture of a single silver bromide crystal exposed to an intense electron beam for some time. At first such crystals are opaque to electrons and are therefore quite uniformly black in such a picture taken with the electron microscope. On being exposed to an electron beam for some time, these crystals begin to show transparent holes and cracks as in this figure. It is thought that this appearance is due to the migration of silver ions within the lattice of the crystal. The appearance of the crystals gradually changes while the exposure to the electron beam is prolonged, but in time a stable state is reached.

B (×30,000) is a picture of very small silver bromide crystals, known as Lippman crystals: not developed or fixed. These would of course be quite invisible in the optical microscope.

C (×25,000) is the picture of a silver bromide crystal exposed to an electron beam, as for the crystal A, and then fixed. This

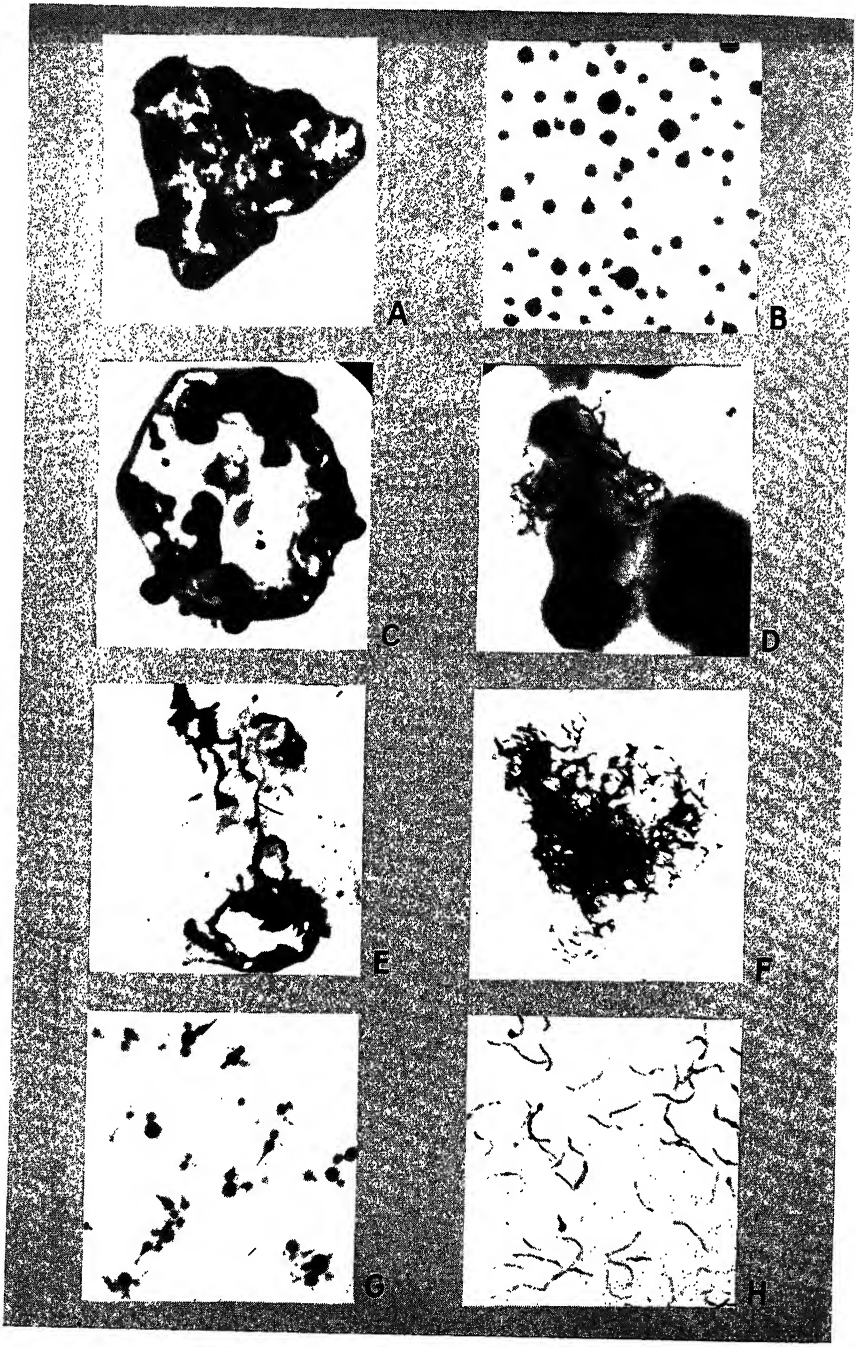
crystal was exposed to an electron beam, while it was observed in the electron microscope. It was then taken from the electron microscope and washed in sodium thiosulphate solution. This dissolved the silver bromide away but left the silver, shown in black. It is to be noted that the reduced silver does not cover the cross-sectional view of the crystal uniformly; the free silver gathers about certain preferred points.

D ($\times$25,000) shows a group of silver bromide crystals exposed to an intense light beam, then put into the developer for a very short time, but not fixed. Some of the silver bromide crystals, the solid black portions of the picture, have not been affected appreciably by the developer. The thread-like structures are filaments of silver, which have been reduced from silver bromide by the developer.

E ($\times$25,000) shows crystals exposed to an intense light beam, then developed longer than those in D, but still only partially developed and then fixed. A faint outline can be seen in the background where the original crystal has left its trace in the gelatin. There are many small particles of silver present, confirming the view that development starts at certain preferred centers, presumably at the latent-image centers.

F ($\times$25,000) is the picture of a silver bromide grain which had been exposed to intense light and then developed on the electron microscope holder. Although this shows the kind of a silver deposit of which a developed silver grain is composed, it is not the same in general appearance as the grains developed on a photographic plate. In the latter the filaments are so closely packed that most of the grain is completely opaque with filamentary structure in evidence only around the edges.

This particular crystal was developed in amidol. With the exception of pure paraphenylenediamine, all of the common chemical developers produce this typical filamentary structure with some secondary differences. Metol and amidol, for example, produce very fine filaments, whereas hydroquinone produces coarse filaments. The extent to which the filaments can wander


A
B
C
D
E
F
G
H

Key to illustration on page 210

A. Silver bromide crystal exposed to intense electron beam (×21,500).

B. Undeveloped Lippmann crystals (×30,000).

C. Silver bromide crystal exposed to electron beam. Fixed (×25,000).

D. Silver bromide crystals. Partially developed. Not fixed (×25,000).

E. Partially developed silver bromide crystal. Fixed (×25,000).

F. A silver grain developed in amidol (×25,000).

G. Undeveloped Lippmann crystals (×30,000).

H. Developed Lippmann crystals (×30,000).

away from the original crystal is influenced by the rigidity of the surrounding gelatin.

G ($\times$30,000) shows Lippmann crystals exposed to light and partially developed but not fixed. The round dots are the silver bromide crystals: it will be noticed that each one has a projection forming.

H ($\times$30,000) shows Lippmann crystals exposed, fully developed and fixed. Each small crystal has grown into a single filament of silver which is much longer and thinner than the original crystal.

These revelations of the electron microscope bid fair to revolutionize the accepted theories of photographic development processes.

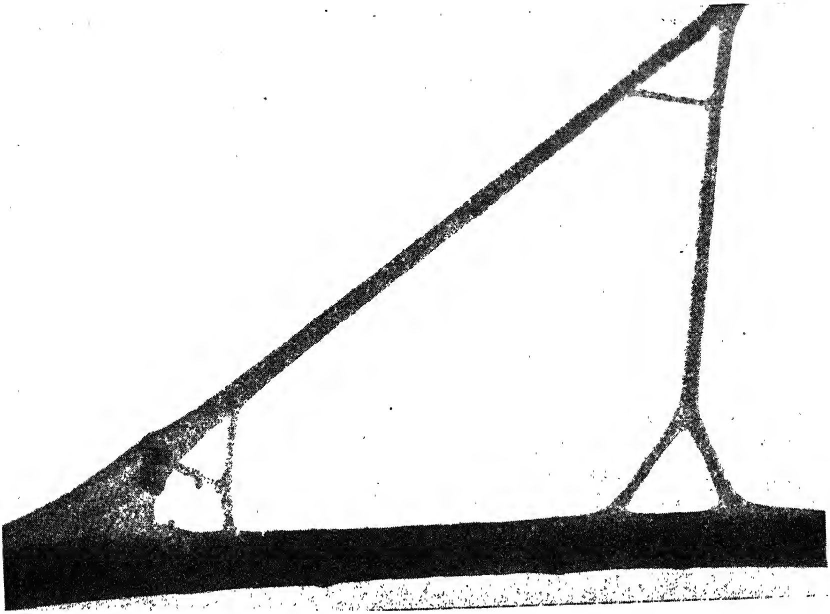
Large Molecules or Molecular Aggregates

Not very much has yet been done on this phase of the usefulness of the electron microscope. The pictures of Cuprene and vinyl chloride need no particular comment at this stage.

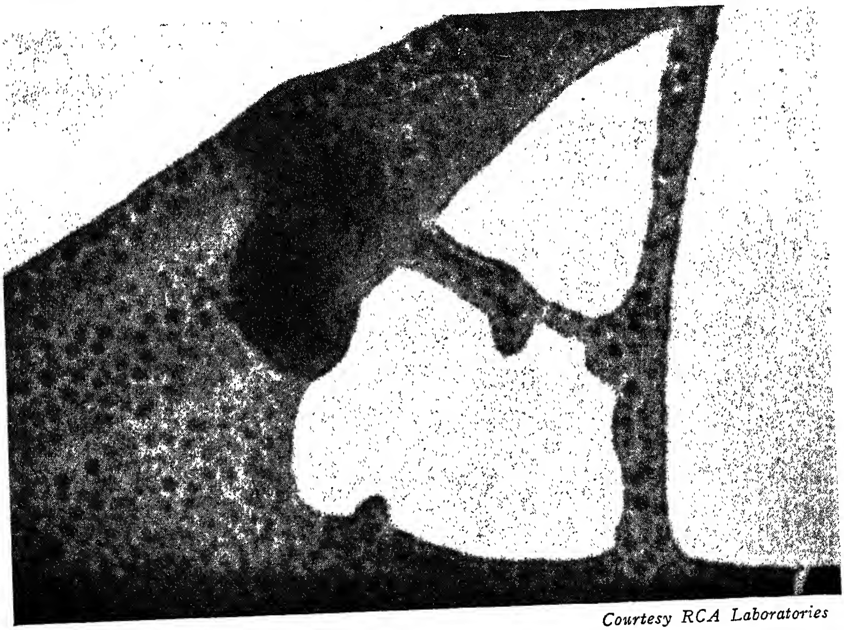
Applications to Biology and Medicine

One can hardly imagine a more important field for the use of the electron microscope than that of biology and medicine. Every one of the pictures shown here is of extreme interest, but very little can be said as yet regarding the interpretation of such revelations as have been made up to the present.

With regard to such pictures as the various types of bacillus and virus, it should be noted that all these samples have to be exposed in a vacuum. In spite of this we have had no indication of any disruption of these structures. It may be that the contents of these bodies are so constructed that the intrinsic pressure of the fluid comes into play and that there are no gas pockets formed within the body to cause explosive disruption of the bacillus or virus membrane. We have not yet determined whether the germs are killed by exposure to the electron beam required for the photography.



Polymerized vinylchloride (×43,000).

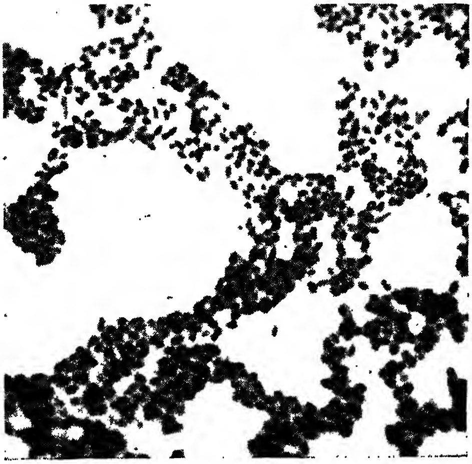


Courtesy RCA Laboratories

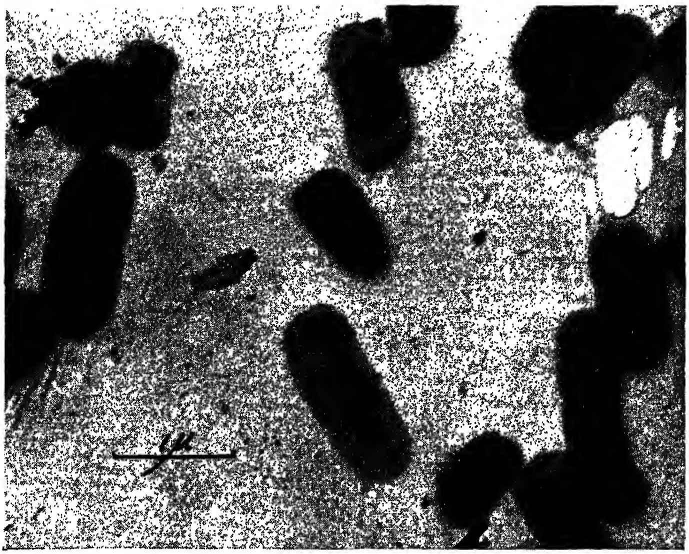
Polymerized vinylchloride (×210,000) (Enlarged photographically).

Cuprene, a polymer of acetylene formed in the presence of copper (×22,700).

Pus cell: Pneumococci around a white blood corpuscle (×8,640). These Pneumococci are similar to those shown in the Frontispiece.

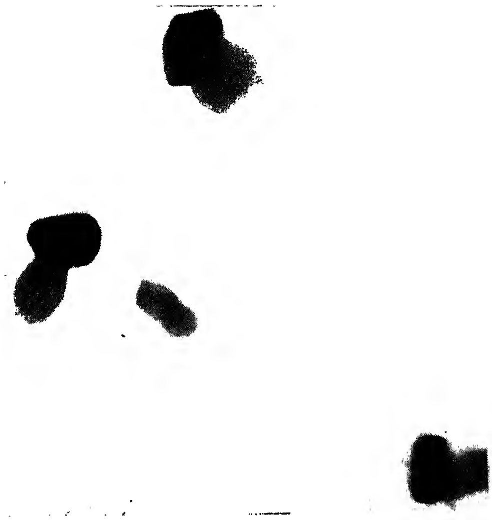


Prodigiosus bacillus under optical microscope (×2,000).



Bacillus Prodigiosus under electron microscope. (×16,400.)

This plate is also designed to compare light microscope and electron microscope pictures of the same specimen—the *Bacillus Prodigiosus*. The optical pictures were taken by Dr. J. Craigie of the Department of Hygiene, University of Toronto. The line marked 1μ gives the scale of the electron picture, this length being 1/10,000th of a cm. or 1/25,000th of an inch. It is quite apparent that the electron microscope may be counted on to show structure in such small bodies. It should be said that none of the electron microscope specimens are stained.

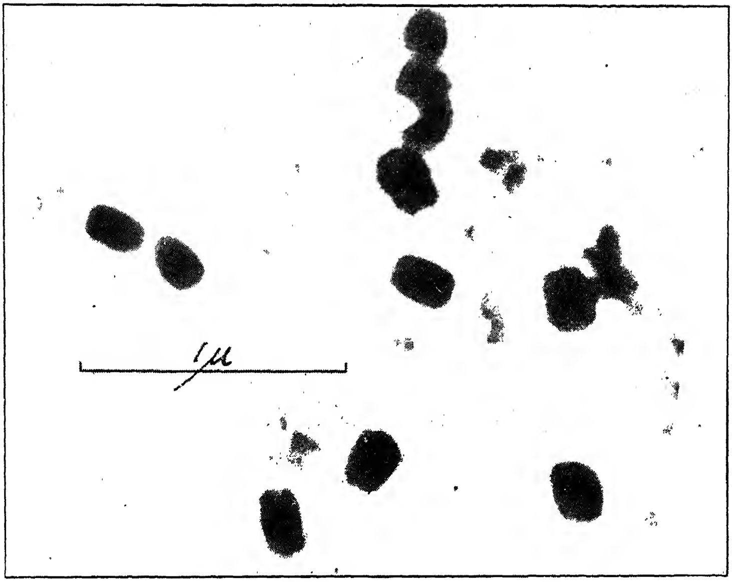


Rickettsia (Typhus) (×13,000). Apparently crystals are formed at edges of bacillus.

Rickettsia (Typhus) (×13,000). Crystals formed at bacillus.

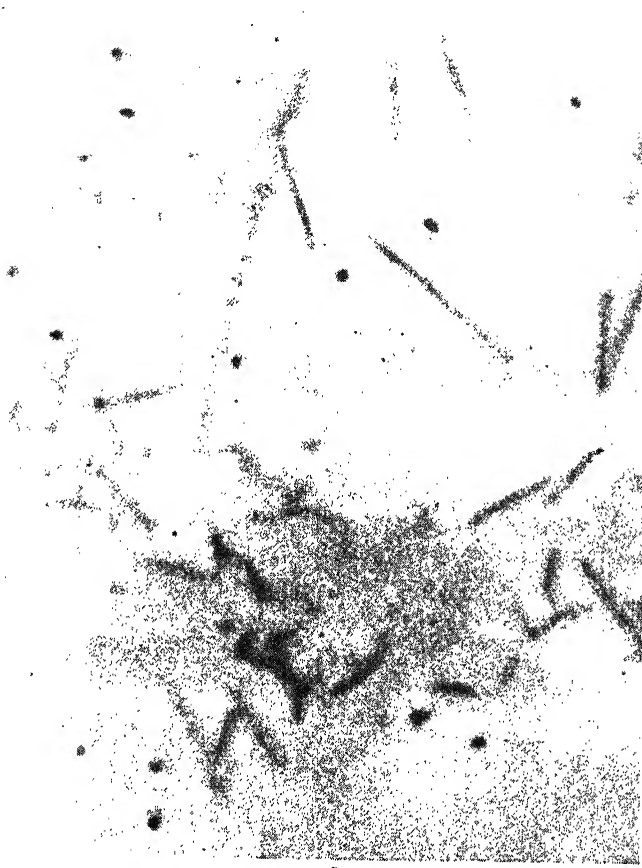
-1μ-

Smallpox virus (×113,400) (Enlarged photographically).



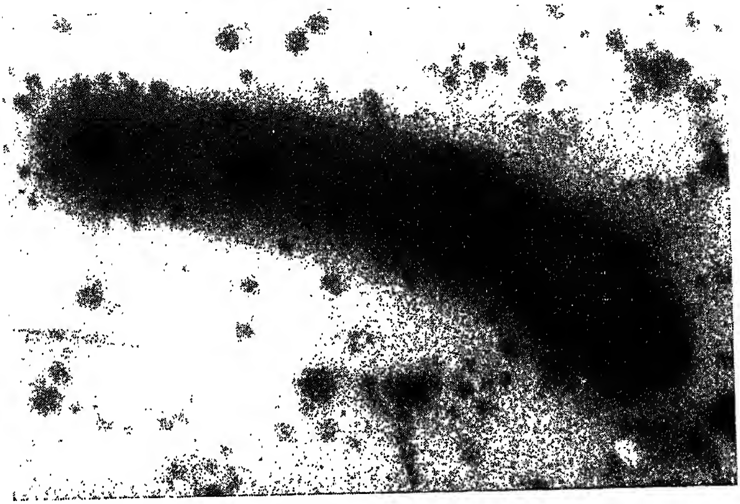
Smallpox viri (×36,000). The barrel shape may be due to centrifuge action; ture visible may or may not be the true structure.

Typhoid germs in presence of bacteriophage (×8,750).



Courtesy of Press Association, Inc.

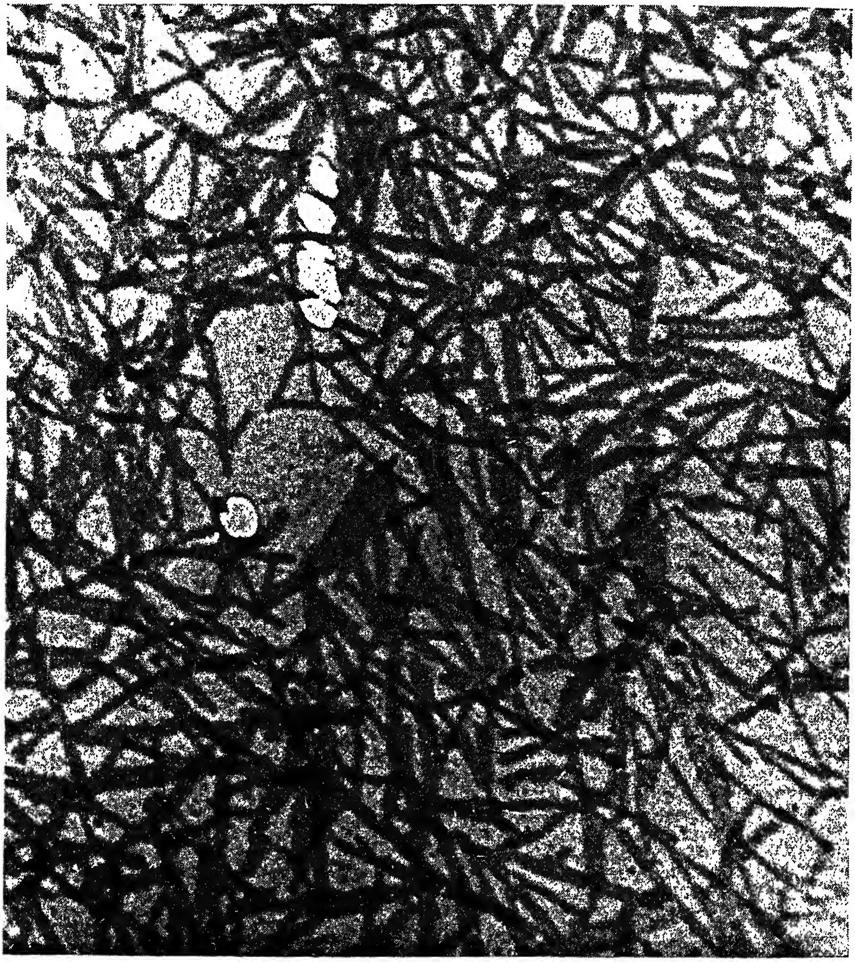
The tiny dots in this picture are influenza virus, each magnified 25,000 times. The long, match-shaped objects are the virus of tobacco mosaic, one of the principal tobacco crop pests (see pages 222 and 223). They were photographed for comparison with the flu germ, which is one of the smallest disease agents known. The picture was made possible by the isolation of the influenza germ by Doctors Leslie A. Chambers and Werner Hene Henle of the Eldridge Reeves Johnson Research Foundation, University of Pennsylvania. The isolation opens the way for the making of a vaccine for influenza.



Tuberculosis bacillus (×31,400).

Tuberculosis bacillus (×26,400).

These two are from the same culture.

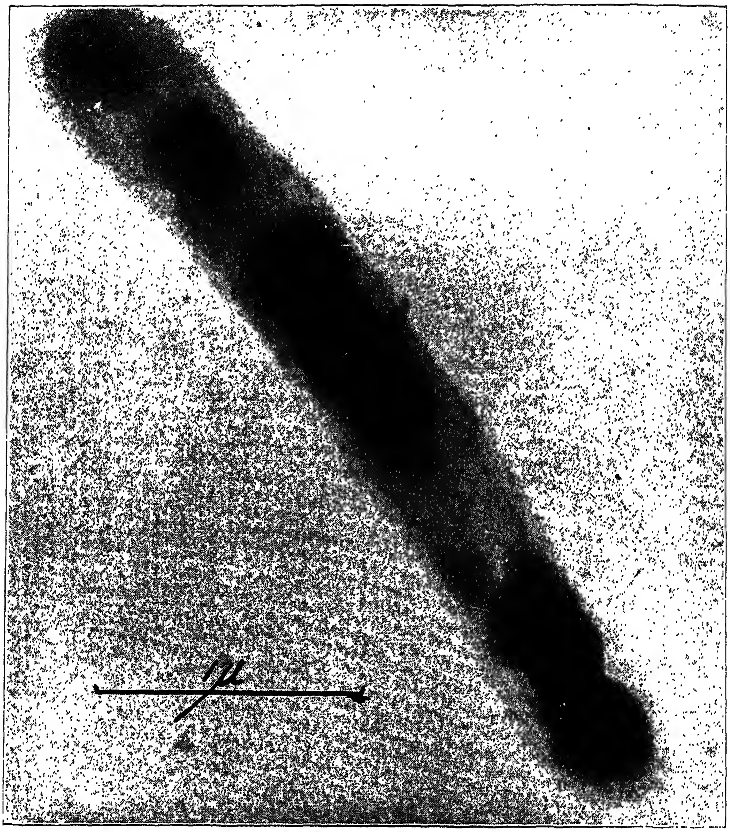


Courtesy RCA Laboratories

Tobacco mosaic (old preparation) (×50,000).

Courtesy RCA Laboratories

Tobacco mosaic (fresh preparation) (×50,000).



Diphtheroid bacillus (×36,300) (stained).

The authors wish to express their gratitude to the publishers for the care and attention given to the production of this book and express the hope that it may accomplish something in establishing the electron microscope in its great role in new experimental technique.

General Bibliography

Technical Monographs

"Geometrische Elektronenoptik," E. Brüche and O. Scherzer, Berlin, Julius Springer, 1934.

Beiträge zur Elektronenoptik, H. Busch and E. Brüche, Leipzig, Barth, 1937.

Television–The Electronics of Image Transmission, V. K. Zworykin and G. A. Morton, New York, J. Wiley & Sons, 1940.

Elektronen–Ubermikroskopie, M. von Ardenne, Berlin, Julius Springer, 1940.

Technical Papers

Burton, E. F., Hillier, J. and Prebus, A.: (1) *Canadian Medical Association Journal*, 42, 116-119 (1940); (2) *Physical Review*, 56, 1171-2 (1939).

Prebus, A. and Hillier, J.: *Canadian Journal of Research*, A17, 49-63 (1939).

Hillier, J.: *Canadian Journal of Research*, A17, 64-69 (1939).

Columbian Colloidal Carbons, Vol. II, New York, Columbian Carbon Co., 1940.

Hall, C. E. and Schoen, A. L.: *Journal of the Optical Society of America*, 31, 281-285 (1941).

Hambly, D. H. and Watson, J. H. L.: *Transactions Royal Society of Canada*, Sec. V, 61 (1941).

R. C. A. Catalogue, "Into Unseen Worlds," 1941.

Mudd, S. and Lackman, D. B.: *Journal of Bacteriology*, 41, 415-420 (1941).

Morton, H. E. and Anderson, T. F.: *Proceedings of the Society for Experimental Biology and Medicine*, 46, 272-276 (1941).

Anderson, T. F. and Stanley, W. M.: *Journal of Biological Chemistry*, 139, 339-344 (1941).

Stanley, W. M. and Anderson, T. F.: *Journal of Biological Chemistry*, 139, 325-338 (1941).

Popular Magazine Articles

The Toronto Saturday Night, Dec. 17, 1938, contains the first reproductions of pictures taken at the University of Toronto.

Maclean's Magazine (Toronto), April 1, 1939, p. 9, by R. E. Beamish.

Scientific American, July, 1940, p. 20, by Jean Harrington.

The Rubber Age, August, 1940, p. 309.

Canadian Chemistry and Process Industries, November, 1940, p. 583.

The Saturday Evening Post, Feb. 7, 1942, p. 27, by J. L. Nicholson.

Index